MENGENAL TEORI PENGUKURAN BERBASIS MODEL RASCH

Wahyu Widhiarso
Fakultas Psikologi Universitas Gadjah Mada
Yogyakarta

Keterbatasan Teori Tes Klasik

PARAMETER BUTIR TERGANTUNG SAMPEL

Parameter butir (tingkat kesulitan atau daya beda) tergantung sampel

Misalnya butir menjadi mudah jika sampelnya orang yang pandai

Istilah mudah dalam skala psikologi dikaitkan dengan endorsability, kecenderungan untuk dijawab YA atau ditanggapi dengan opsi SANGAT SETUJU/SANGAT SESUAI

Keterbatasan Teori Tes Klasik

TINGGI RENDAHNYA KEMAMPUAN PESERTA TERGANTUNG DARI TINGKAT KESULITAN TES

- Peserta tes akan kelihatan lebih mampu jika mendapatkan banyak jawaban benar dari tes yang mudah

Skala Kecakepan Subjektif

Versi A

Pernyataan	Ya	Tidak
Saya memiliki kecakepan dalam tingkat lumayan		
Dibandingkan dengan orang sekampung, saya terlihat cakep		
Dilihat dari kejauhan, saya terlihat cakep		

Versi B

Pernyataan	Ya	Tidak
Tidak terbantahkan bahwa saya adalah orang yang cakep		
Saya lebih cakep dibandingkan dengan para cover boy		
Dipandang dari dekat atau jauh, saya kelihatan cakep		

Berikut ini adalah contoh skala yang mengukur kecakapan subjektif. ada dua versi yaitu versi A dan versi B. Dilihat dari konten butirnya terlihat bahwa butir-butir skala Versi A relatif memiliki tingkat kesulitan yang tinggi dibanding dengan Versi B. Akibatnya, skor yang sama dari 2 tes yang berbeda memiliki ketidaksetaraan dalam menginterpretasikan kecakepan.

SUBJEK 1 yang mendapatkan skor 3 di Skala Versi A

tidak setara dengan

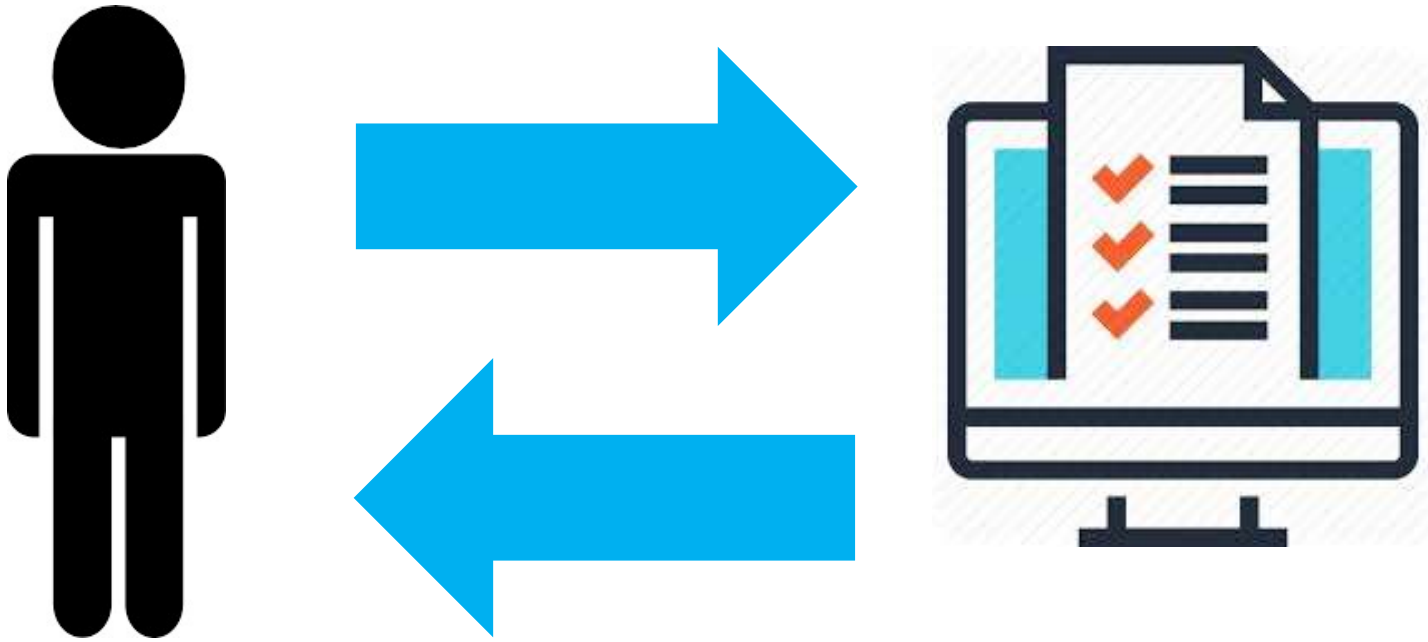
SUBJEK 2 yang mendapatkan skor 3 pada Skala Versi B.

TINGGI RENDAHNYA KEMAMPUAN PESERTA TERGANTUNG DARI TINGKAT KESULITAN TES

Inseparability of Person and Item Parameters

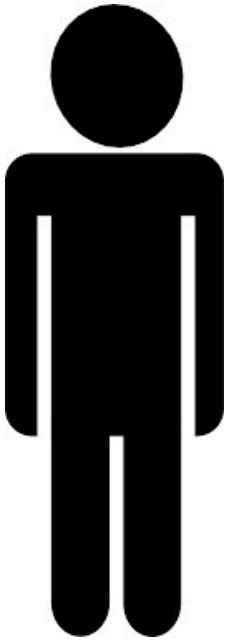
Fenomena ini dinamakan dengan Inseparability of Person and Item Parameters. Menunjukkan bahwa kemampuan orang dan parameter butir merupakan satu kesatuan yang tidak dapat dipisahkan.

Hal ini membuat pengukuran menjadi subjektif. Oleh karena itu diperlukan sebuah teknik analisis yang membuat pengukuran menjadi objektif dengan cara memisahkan antara kemampuan orang karakteristik tes.



ORANG **TES**

Sebuah tes yang objektif harus terpisah dari orang yang diukur.



ORANG

TES

Tes yang objektif adalah tes berdiri secara terpisah dengan objek yang di tes sehingga memenuhi sifat independen

MODEL RASCH

- Seorang ahli matematika Georg Rasch memiliki konsep untuk mengatasi masalah ini. Ia kemudian bertemu dengan Benjamin Wright dari Universitas Chicago (AS) dan mengembangan bersama model ini.
- Salah satu hal yang mereka diskusikan adalah bagaimana pengukuran di dalam bidang ilmu sosial memiliki unit pengukuran yang sama halnya dengan unit ukut di dalam ilmu fisika, misalnya meter atau gram.

MODEL RASCH

- Mereka kemudian menemukan unit yang dinamakan dengan LOGIT yang merupakan hasil dari penerapan FUNGSI LOGARITMA pada ODDS RATIO
- ODDS adalah rasio antara probablilitas orang menjawab benar sebuah butir dengan probabilitas menjawab salam

$$Odds = \frac{\Pr}{1 - \Pr}$$

$$\Pr = \frac{Odds}{1 + Odds}$$

ODDS RATIO

- Mereka kemudian menemukan unit yang dinamakan dengan LOGIT yang merupakan hasil dari penerapan FUNGSI LOGARITMA pada ODDS RATIO
- ODDS adalah rasio antara probablilitas orang menjawab benar dengan probabilitas menjawab salah

Misalnya sebuah butir dijawab dengan benar oleh 40 orang dan dijawab salah oleh 60 orang maka ODDS RATIO dari hal ini adalah:

ODDS RATIO =40/60 = 0,67

Jika soal tersebut dijawab benar oleh separuh peserta tes maka ODDS RATIO = 1

LOGIT SCALE

Jumlah Jawaban Benar	10	20	40	50	60	80	90
Jumlah Jawaban Salah	90	80	60	50	40	20	10
ODDS RATIO	0.11	0.25	0.67	1	1.5	4	9
LN(ODDS RATIO)	-2.20	-1.39	-0.41	0.00	0.41	1.39	2.20

Setelah ODDS RATIO di transformasi dengan menggunakan fungsi logaritma natural maka didapatkan kontinum skor yang bergerak mirip dengan distribusi normal.
Bergerak antara sekitar -3 hingga 3.
Hasil transformasi ini dinamakan dengan SKALA LOGIT

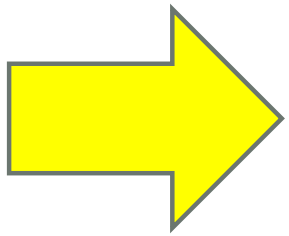
Berhasil	Gagal	Probabiltas	ODDS	LOGIT
99	1	0.99	99.00	4.60
95	5	0.95	19.00	2.94
90	10	0.90	9.00	2.20
80	20	0.80	4.00	1.39
70	30	0.70	2.33	0.85
60	40	0.60	1.50	0.41
50	50	0.50	1.00	0.00
40	60	0.40	0.67	-0.41
30	70	0.30	0.43	-0.85
20	80	0.20	0.25	-1.39
10	90	0.10	0.11	-2.20
5	95	0.05	0.05	-2.94
1	99	0.01	0.01	-4.60

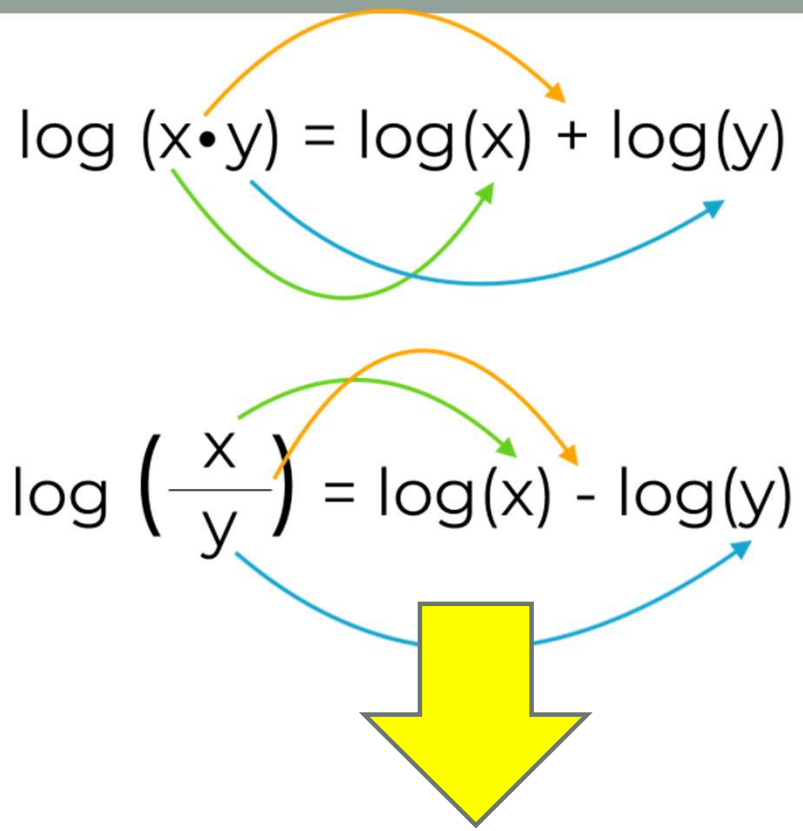
Terlihat bahwa jarak antara skor satu dengan skor lainnya bersifat equal interval

Konversi logit membuat data yang awalnya bersifat cacah (COUNT) menjadi data yang memiliki metrik sehingga bisa dikalkulasi (SUM)

$$Odds = \frac{\Pr}{1-\Pr}$$

$$\Pr = \frac{Odds}{1+Odds}$$





Gambar ini menunjukkan bahwa ketika di masih menjadi logaritma yang awalnya adalah sebuah **pembagian** berubah menjadi **pengurangan**. Hal ini dikarenakan logaritma membuat fungsi pembagian menjadi pengurangan.

$$P\,|\,(x=1) = \frac{e^{b-d}}{1+e^{b-d}}$$

$$P = \frac{Odds}{1 + Odds}$$

$$= \frac{P/(P-1)}{1 + P/(P-1)}$$

----- fungsi log ----

$$P = \frac{e^{b-d}}{1 + e^{b-d}}$$

Dari rumus kemudian diinterpretasikan bahwa proporsi jawaban benar sebuah butir (P) merepresentasikan kemampuan individu (b)

...sedangkan proporsi menjawab salah (P-1) adalah tingkat kesulitan butir (d).

P (atau probabilitas menjawab butir dengan benar) sama dengan 0,5 ketika kemampuan individu dan tingkat kesulitan butir adalah sama.
P = 1/1+1 = 0,5 (catatan exp(0) sama dengan 1)

TINGKAT KESULITAN BUTIR

=

KEMAMPUAN INDIVIDU

Gambar ini adalah kontinum probabilitas untuk menjawab Benar suatu butir, yang bergerak mulai dari 0 hingga 1. Semakin tinggi, semakin besar peserta tes untuk menjawab benar. Titik tengah dari kontinum ini adalah 0,5 yang menunjukkan bahwa individu memiliki peluang sebesar 50% untuk menjawab benar.

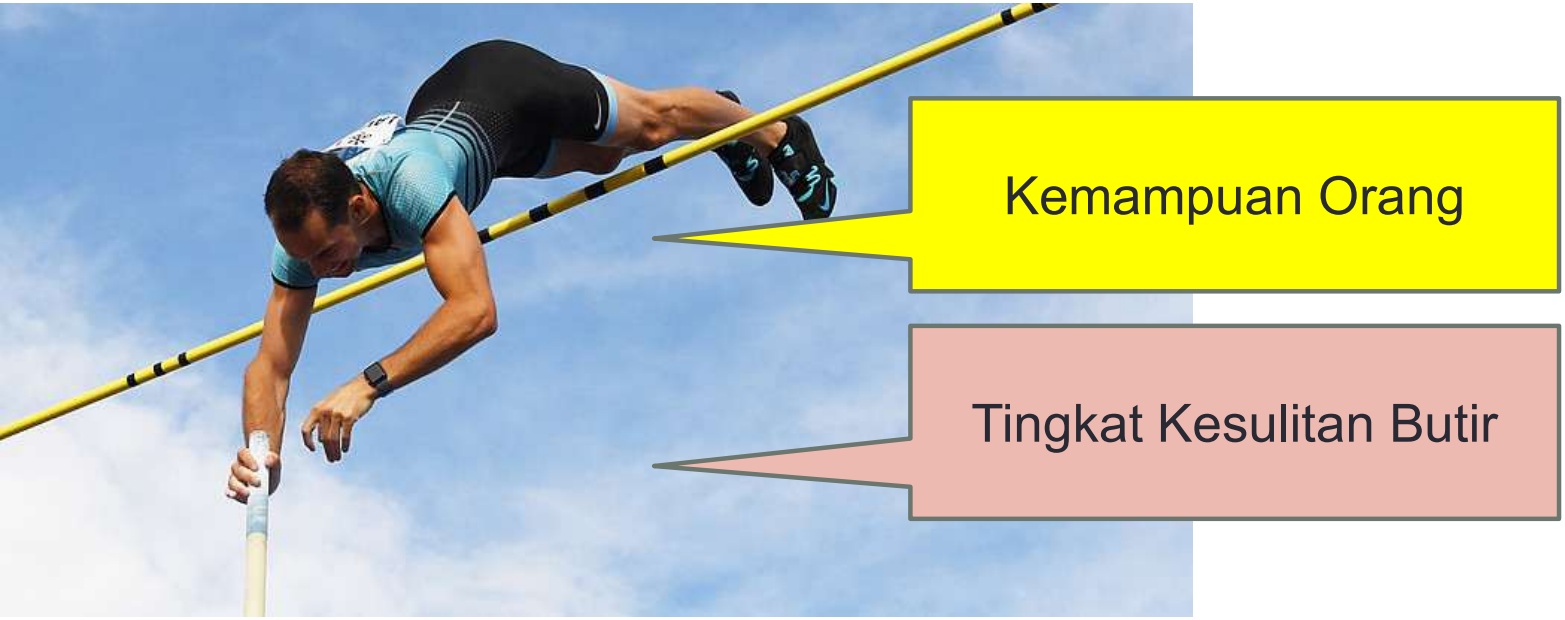
Berdasarkan rumus tadi didapatkan informasi bahwa probabilitas sebesar 0,5 atau 50% didapatkan oleh individu yang memiliki kemampuan setara dengan tingkat kesulitan butir.

Pada gambar ini terlihat bahwa kemampuan orang terletak di bawah tingkat kesulitan butir sehingga probabilitas untuk berhasilnya rendah (< 0.50)

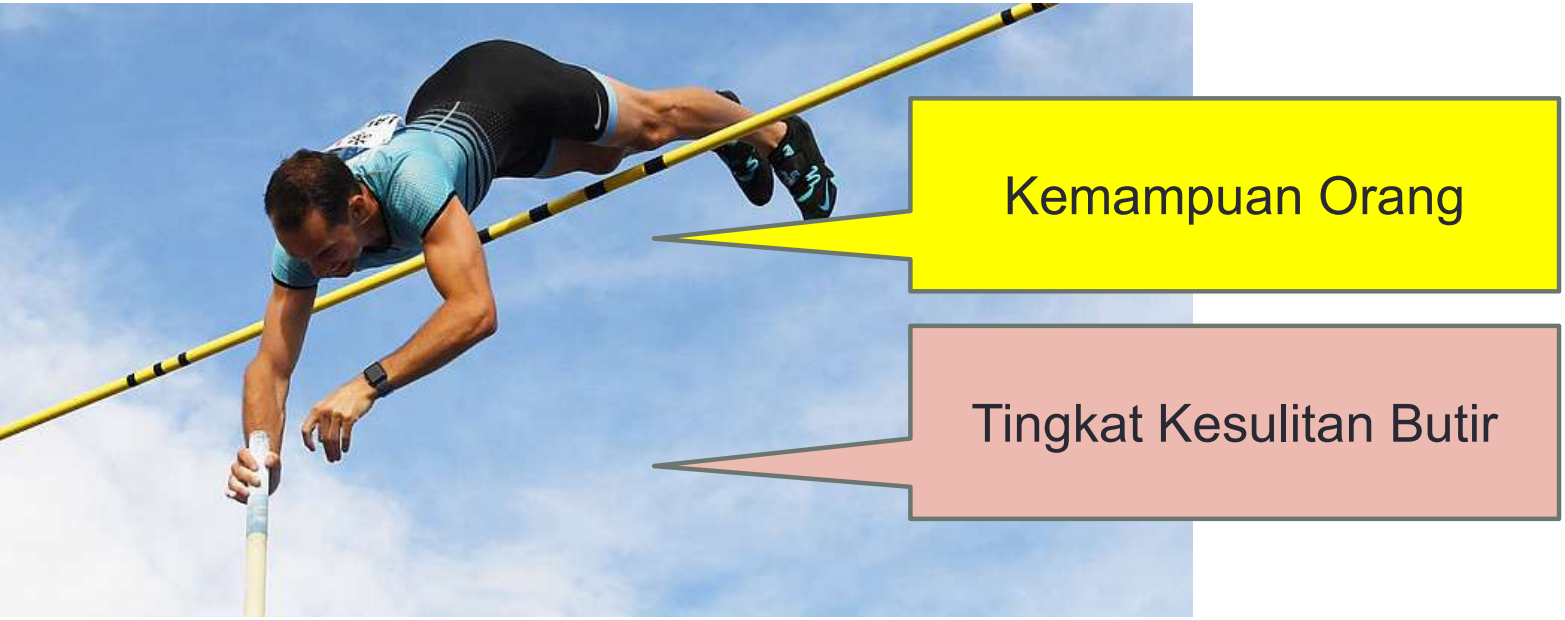
Tingkat Kesulitan Butir

Kemampuan Orang

Pada gambar ini terlihat bahwa kemampuan orang lebih tinggi daripada dengan tingkat kesulitan butir sehingga probabilitas untuk menjawab benar cukup tinggi (> 0.50)



Pada gambar ini terlihat bahwa kemampuan orang lebih tinggi daripada dengan tingkat kesulitan butir sehingga probabilitas untuk menjawab benar cukup tinggi (> 0.50)



- SS

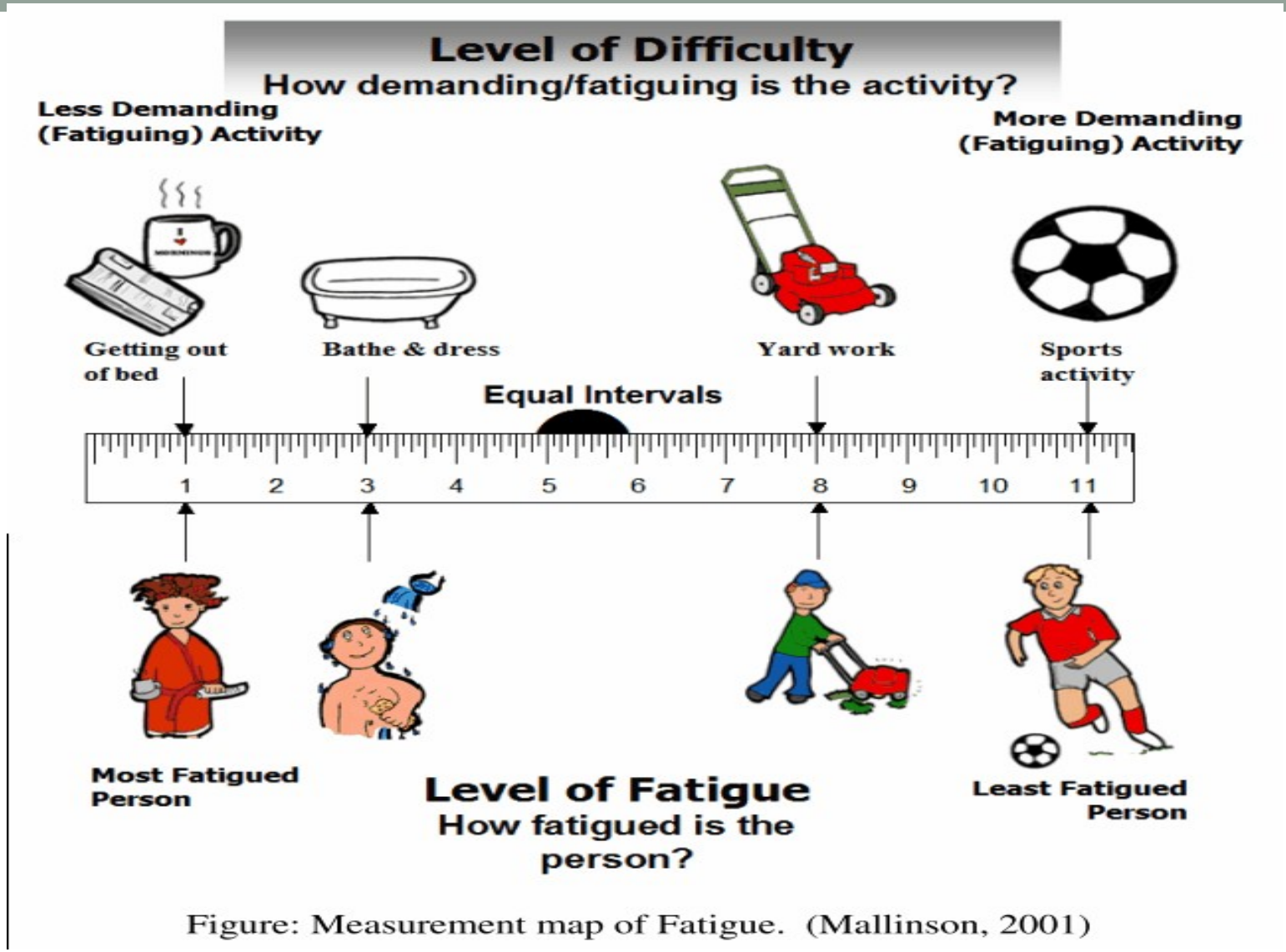
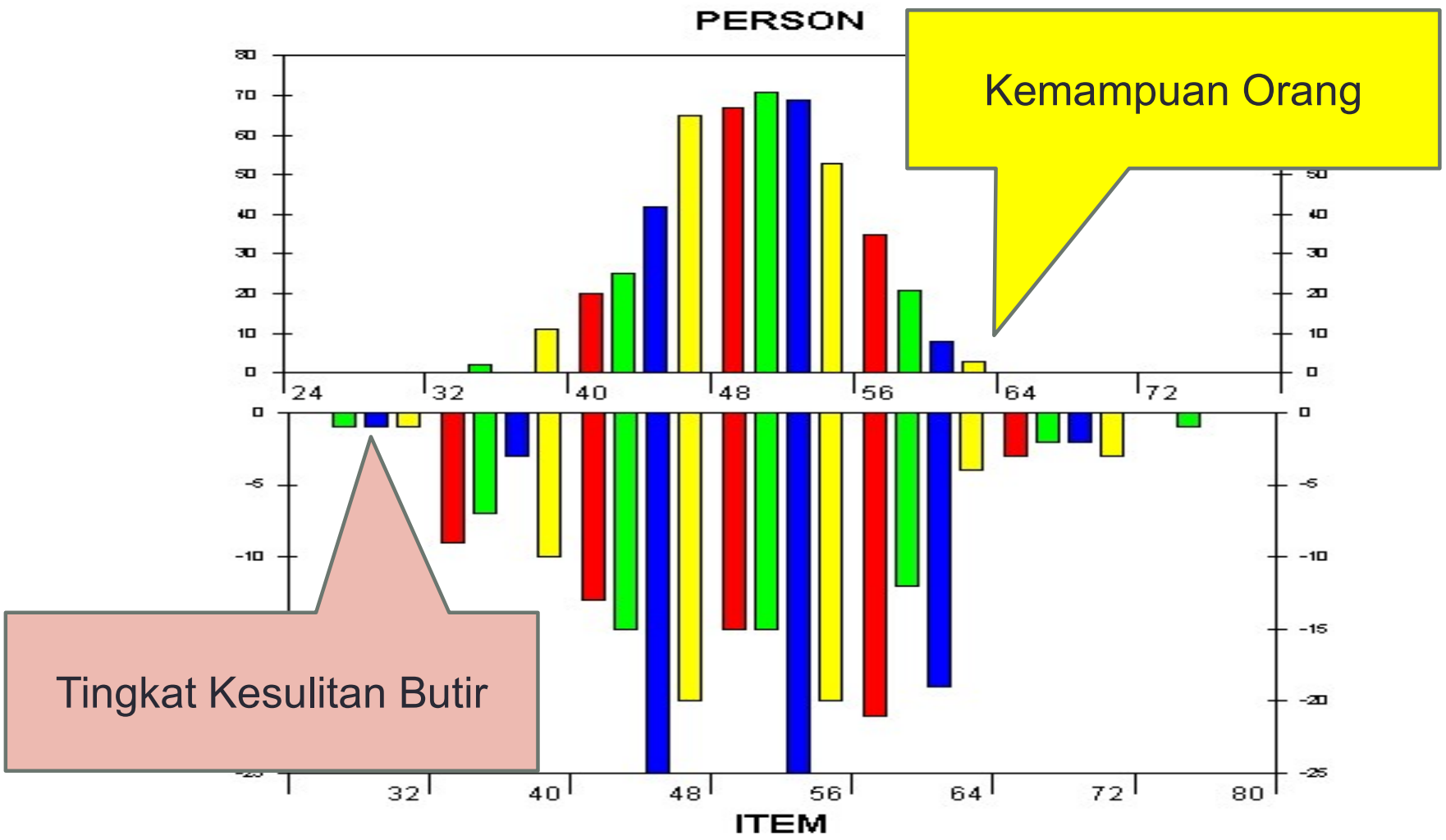
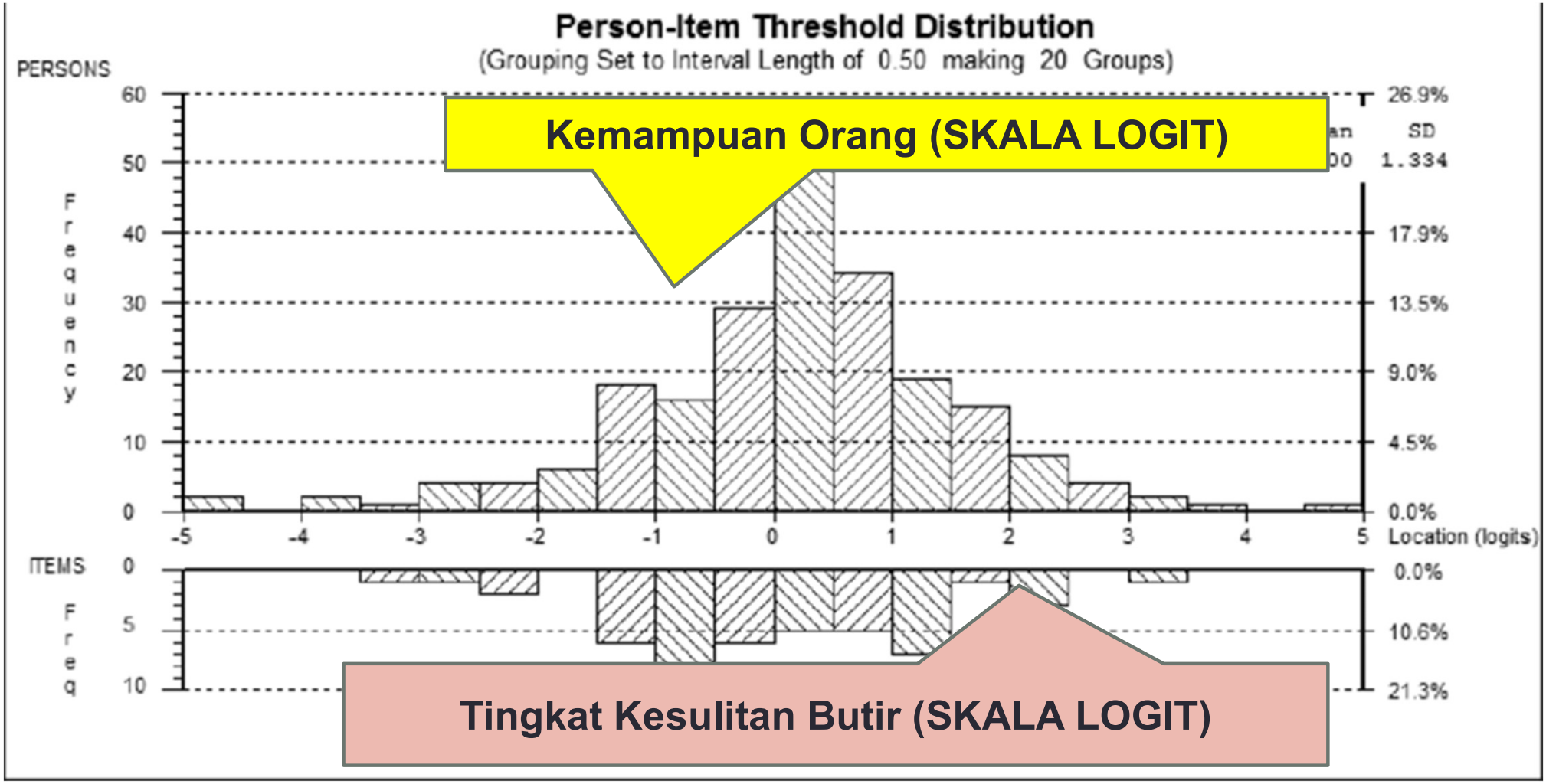


Figure: Measurement map of Fatigue. (Mallinson, 2001)


PERSON
Kemampuan Orang
Tingkat Kesulitan Butir
ITEM


Person-Item Threshold Distribution
(Grouping Set to Interval Length of 0.50 making 20 Groups)
PERSONS
Frequency
60
50
40
30
20
10
0
26.9%
17.9%
13.5%
9.0%
4.5%
0.0%
SD
1.334
-5
-4
-3
-2
-1
0
1
2
3
4
5
Location (logits)
ITEMS
Freq
0
5
10
0.0%
10.6%
21.3%
Kemampuan Orang (SKALA LOGIT)
Tingkat Kesulitan Butir (SKALA LOGIT)

Person-Item Threshold Distribution
(Grouping Set to Interval Length of 0.50 making 20 Groups)
PERSONS
Frequency
60
50
40
30
20
10
0
26.9%
4.5%
0.0%
-5
-4
-3
-2
-1
0
1
2
3
4
5
Location (logits)
ITEMS
Freq
0
5
10
0.0%
10.6%
21.3%
Kemampuan orang dan
Tingkat kesulitan dalam
Skala atau Metrik yang sama

Tingkat Kemampuan Orang

Tingkat Kesulitan Tes

Tingkat Kemampuan Orang

Tes Produktif Karena Ada INFORMASI

Tingkat Kesulitan Tes

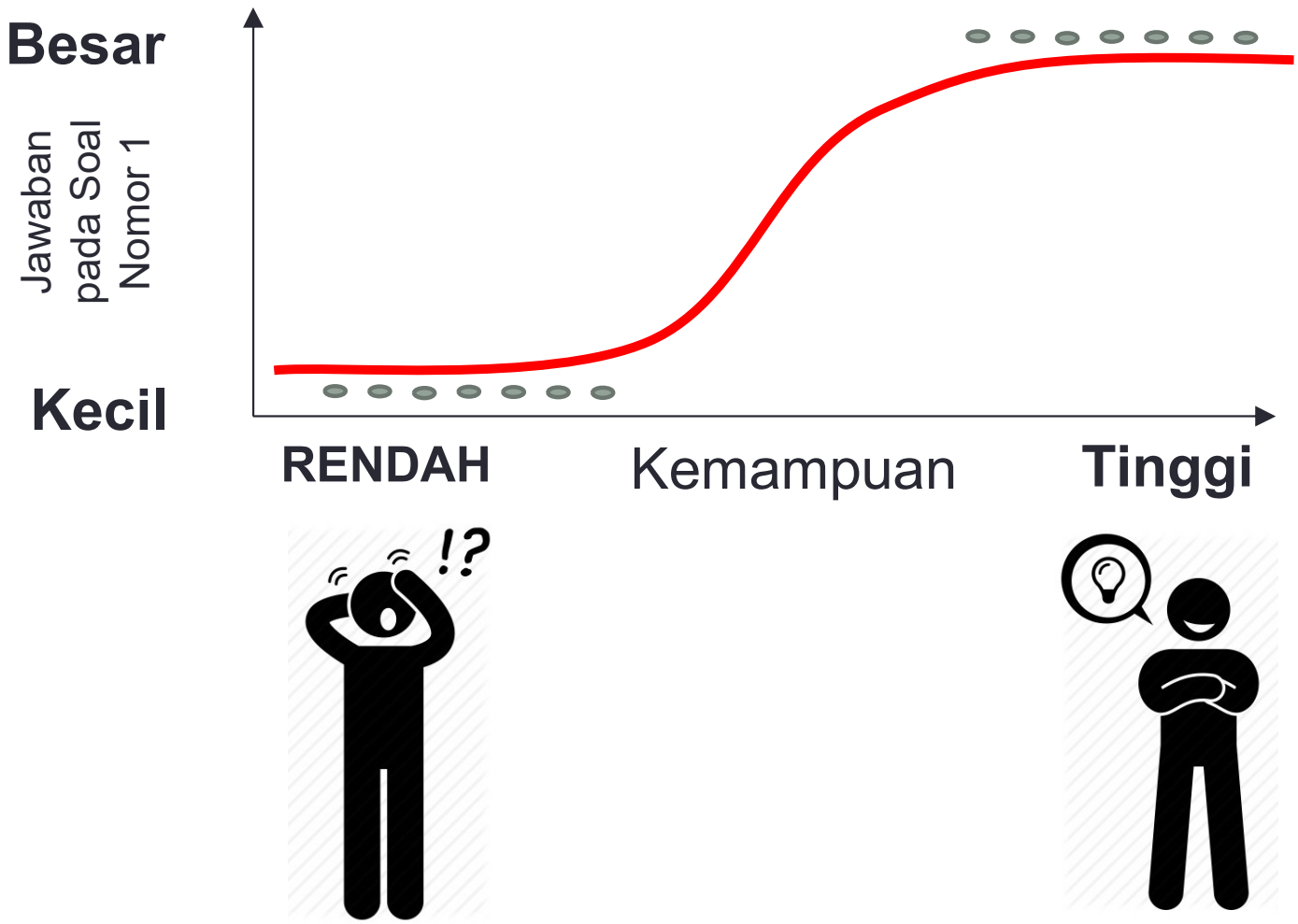
Tes atau skala yang dikembangkan dengan menggunakan basis Model Rasch diharapkan memiliki butir dengan tingkat kesulitan yang bervariasi mulai dari mudah hingga sulit. demikian juga dengan skala psikologi diharapkan memiliki butir dengan konten ukur yang untuk di setujui hingga butir yang memiliki konten ukur sulit untuk disetujui. Tujuannya adalah untuk membuat alat ukur penggaris. Filosofi pengembangan tes berbasis model ini adalah yang dikembangkan oleh thurstone dan Guttman.

$$P\,|\,(x=1) = \frac{e^{b-d}}{1+e^{b-d}}$$

Ini adalah rumus dari pengukuran berbasis model Rasch. Jika tingkat kesulitan (b) dibuat konstan, misalnya sebesar 0 dalam skala logit dan kemampuan individu dibuat bervariasi mulai dari -4 hingga 4 maka didapatkan grafik seperti ini.

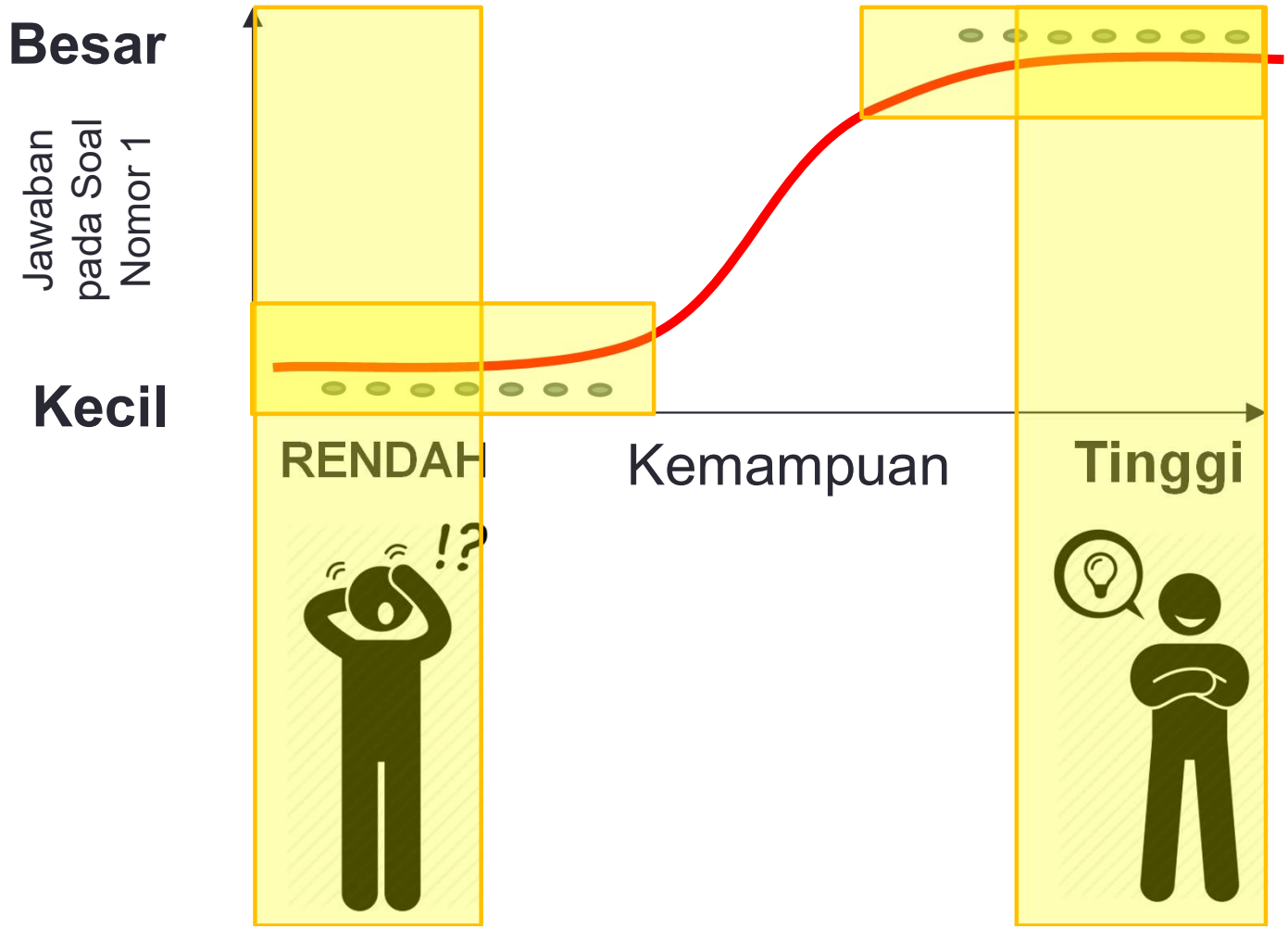
Arti Gambar Kurva

Gambar ini menunjukkan bahwa Orang yang memiliki kemampuan lebih tinggi akan memiliki peluang lebih besar dalam menjawab benar dibanding dengan orang yang kemampuannya rendah



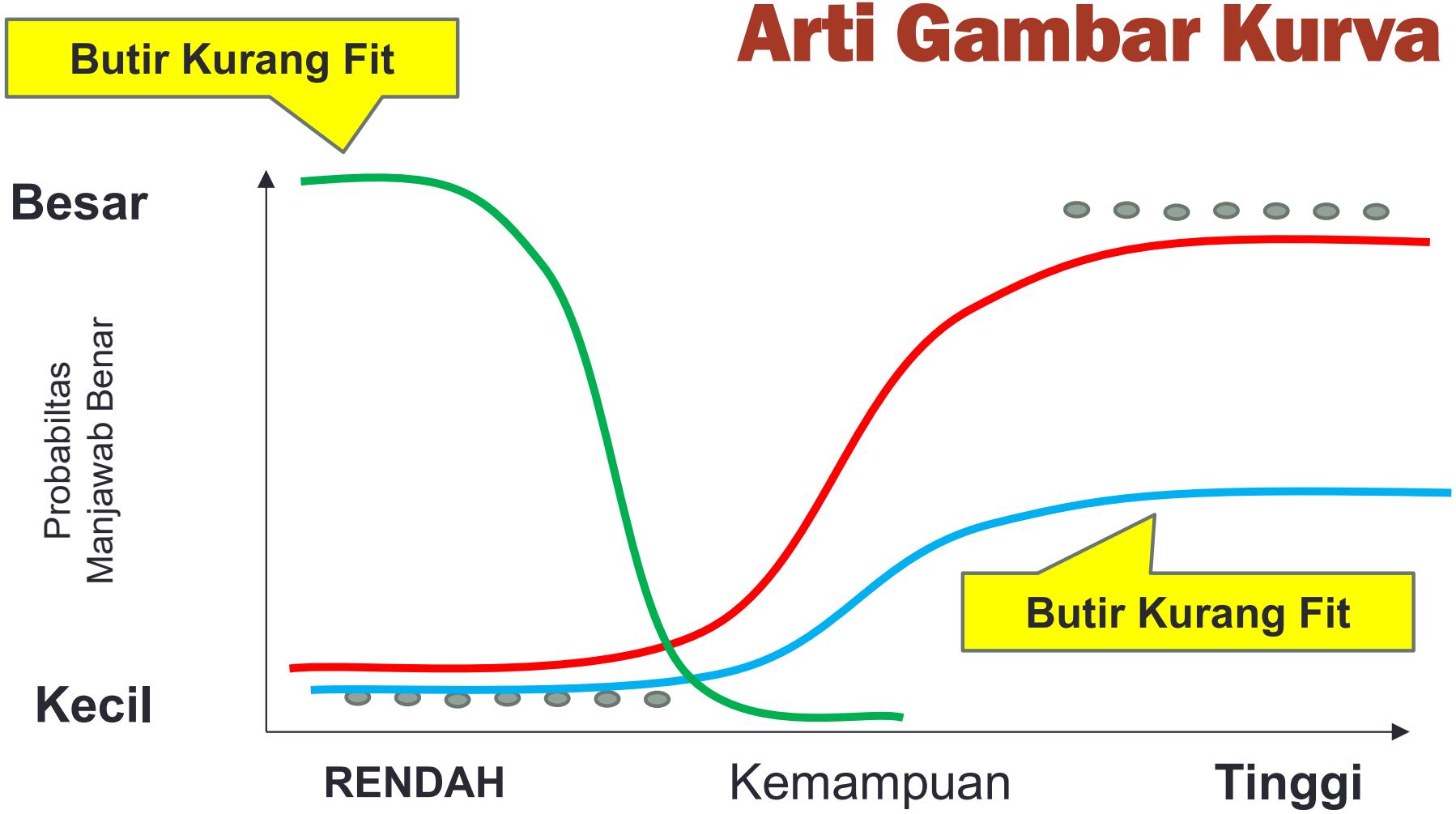
Arti Gambar Kurva

Gambar ini menunjukkan bahwa Orang yang memiliki kemampuan lebih tinggi akan memiliki peluang lebih besar dalam menjawab benar dibanding dengan orang yang kemampuannya rendah

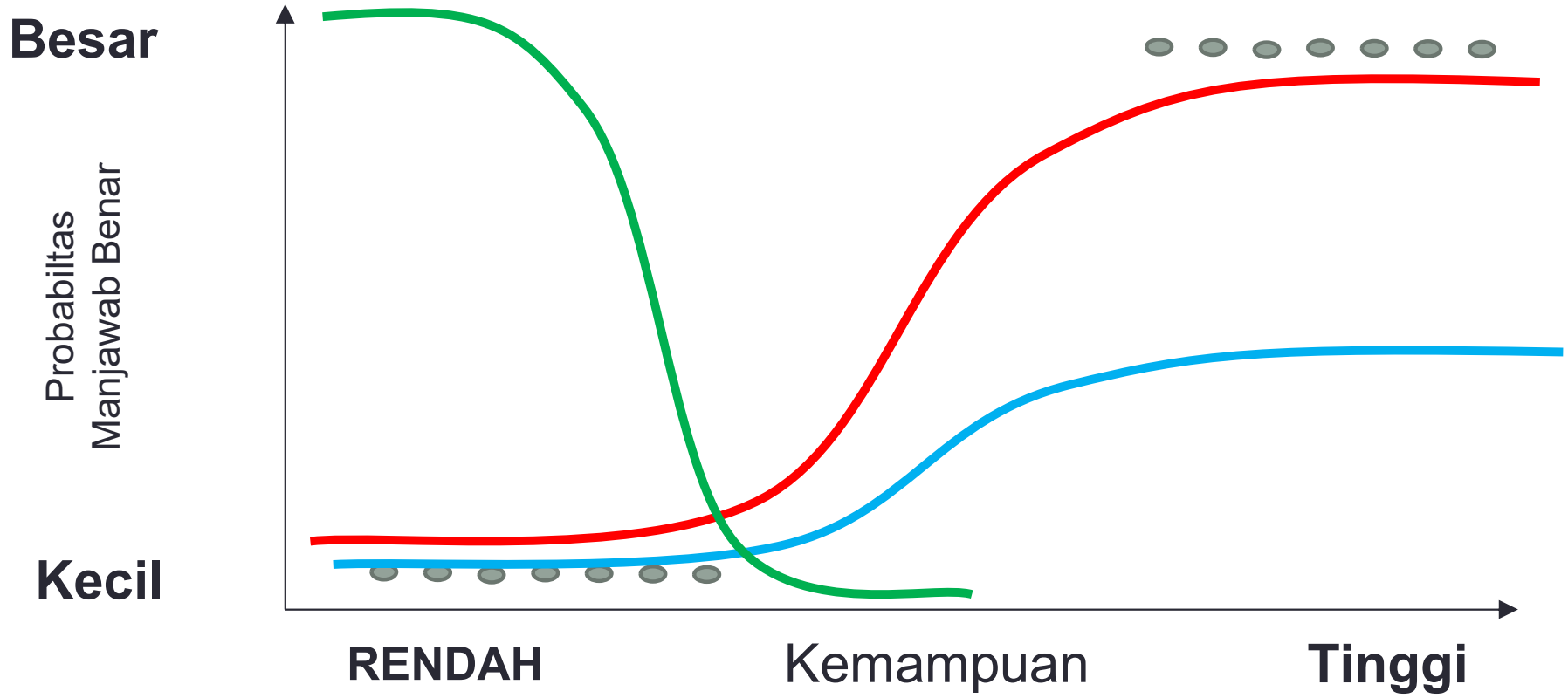


Di dalam analisis berbasis model Rasch, baik buruknya performansi sebuah bujur dilihat dari kesesuaian kesesuaiannya dengan grafik yang mengikuti Model Rasch. Butir-butir yang tidak sesuai dengan grafik model Rasch. Butir-butir yang tidak dengan model Rasch dapat dikatakan butir yang tidak fit atau tidak dengan model. Butir-butir ini merupakan butir kandidat untuk dihapus dari tes di versi akhir.

Arti Gambar Kurva



Arti Gambar Kurva

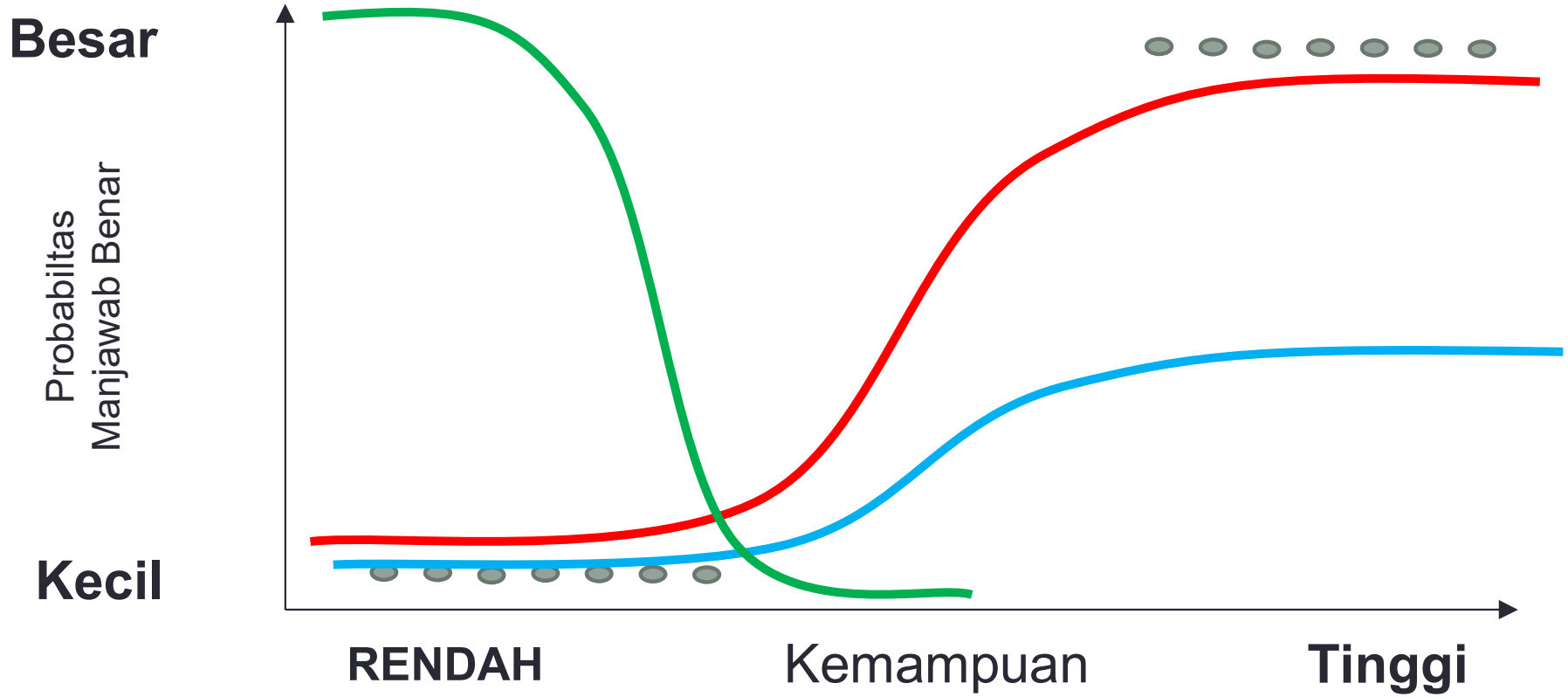


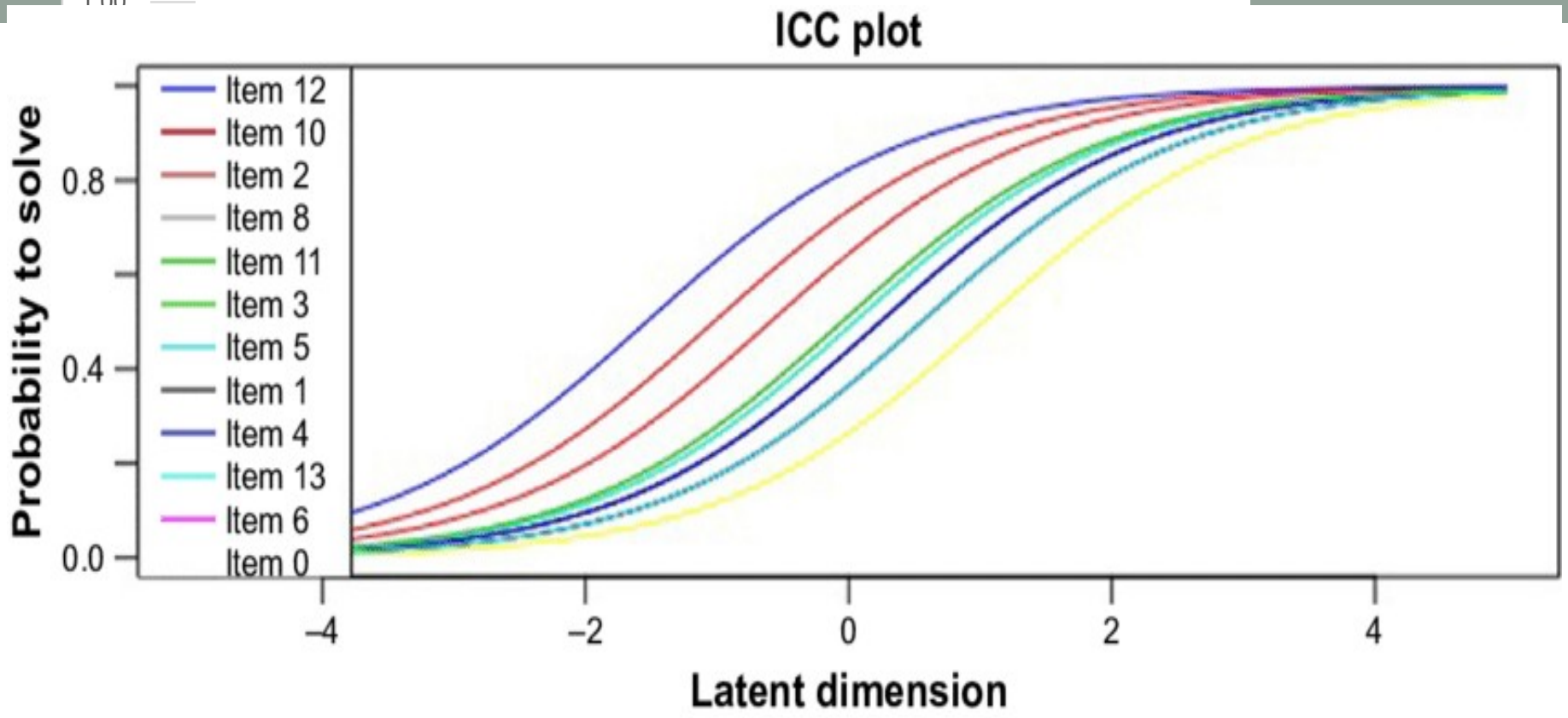
MENGENAL TEORI PENGUKURAN BERBASIS MODEL RASCH (BAGIAN 2)

Wahyu Widhiarso
Fakultas Psikologi Universitas Gadjah Mada

Pada video terakhir kita sudah belajar mengenai prosedur untuk mengevaluasi performansi butir berdasarkan model Rasch, yaitu dengan cara **menguji ketepatan model sebuah butir berdasarkan model Rasch**. Butir yang tidak sesuai dengan model Rasch akan memiliki nilai ketepatan model yang rendah.

Evaluasi terhadap ketepatan butir tidak dapat dilakukan dengan menggunakan grafik karakteristik kurva seperti yang terlihat pada gambar ini. Hal ini dikarenakan model RASCH menetapkan bahwa butir di dalam sebuah tes atau skala memiliki **daya beda yang sama**, **yang berbeda hanyalah kesulitan butir**. Terlihat bahwa daya beda pada ketiga butir di bawah ini berbeda-beda. Butir hijau daya bedanya negatif, butir biru daya bedanya rendah dan butir merah daya bedanya tinggi.





Semua butir yang dianalisis oleh model Rasch jika ditampilkan dalam kurva karakteristik butir akan ditampilkan seperti pada gambar di atas. Semua butir memiliki daya beda yang sama, yang ditunjukkan dengan kemiringan garis yang sama

Hal ini dapat terjadi karena di dalam persamaan model Rasch yang kita pelajari kemarin variasi hanya terjadi ada dua koefisien yaitu tingkat kesulitan butir (B) dan kemampuan individu (D).

Tingkat Kesulitan Butir

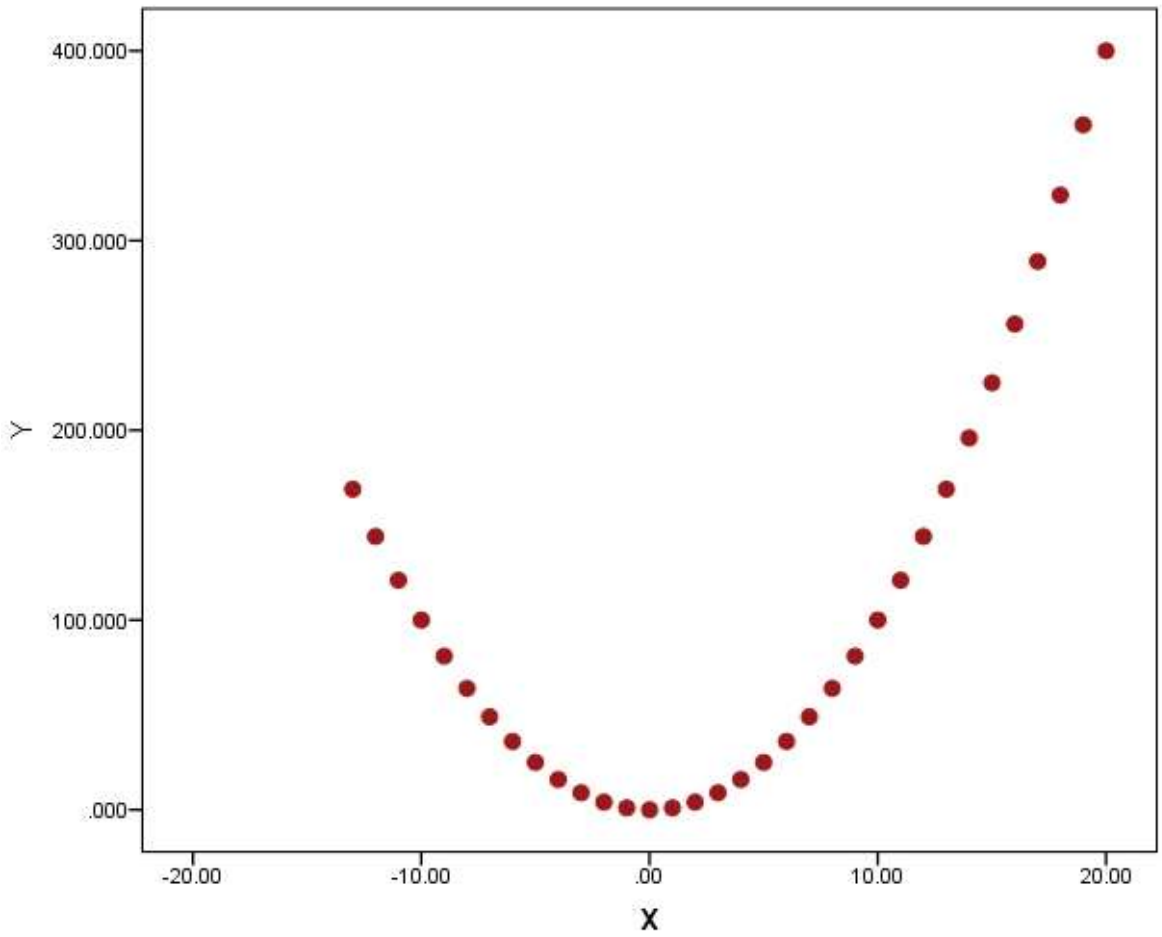
Tingkat Kemampuan Individu

$$P = \frac{e^{b-d}}{1+e^{b-d}}$$

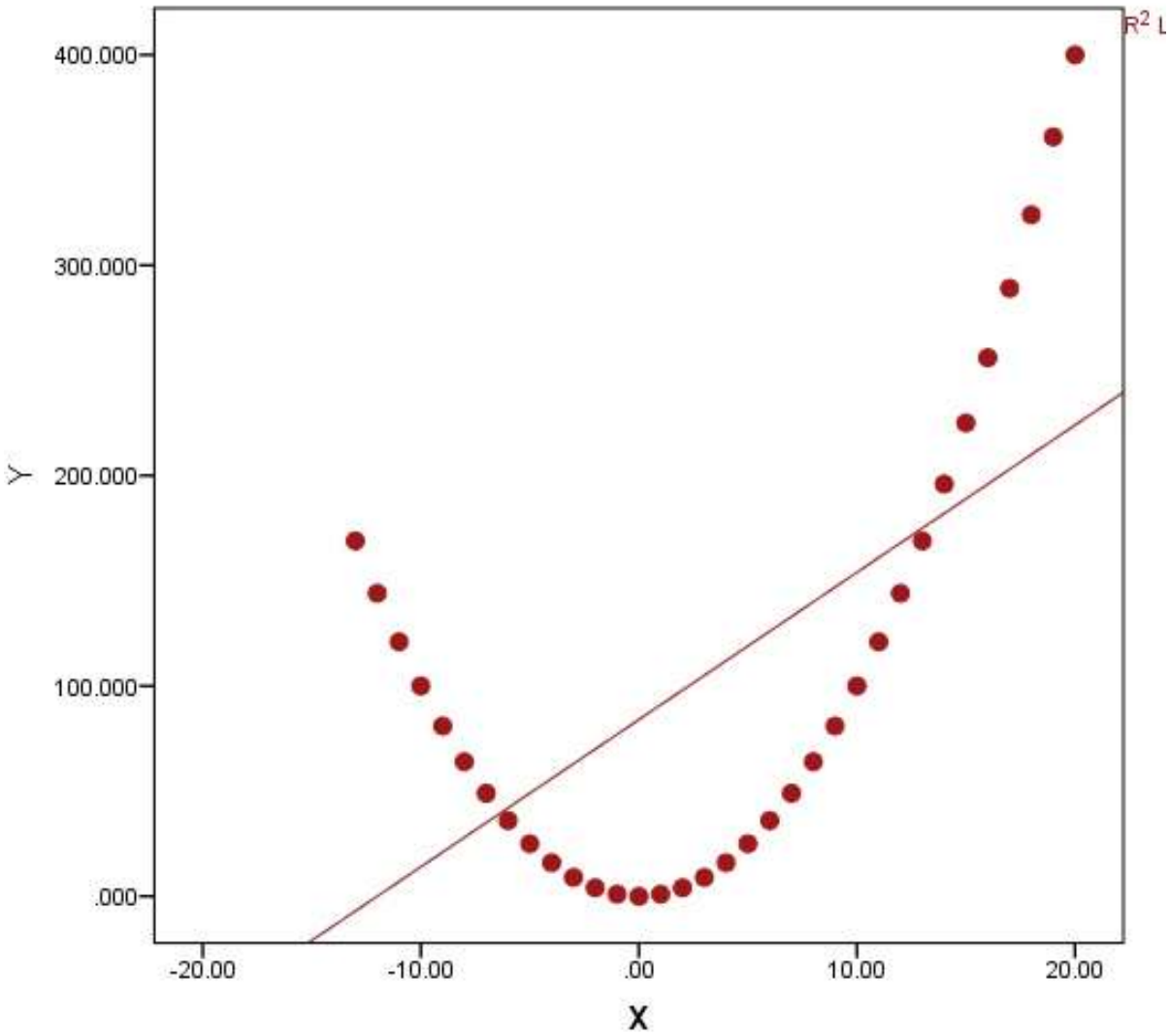
Tidak ada Koefisien Daya Beda butir di dalam Model ini

Kondisi ini sama seperti dengan memprediksi garis ini dengan menggunakan model linier.

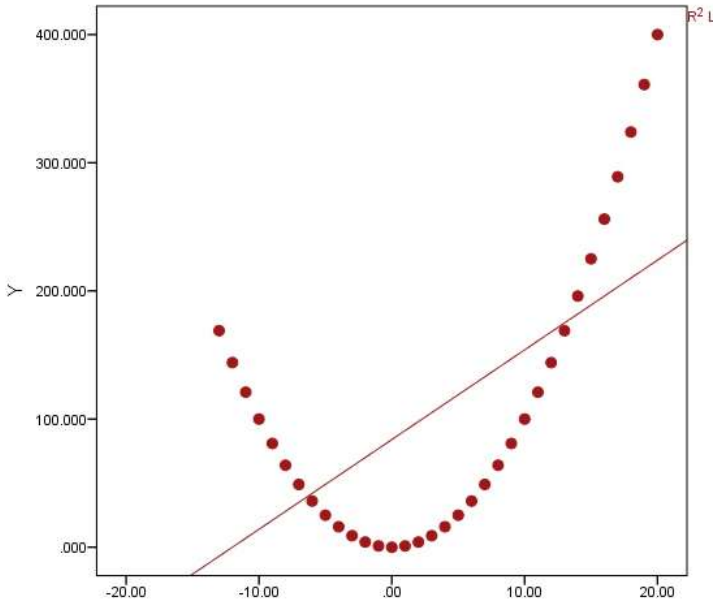
$Y = A + BX$



Hasil analisis memunculkan **garis yang berbentuk lurus** karena memang model yang dipakai adalah model linier



Padahal jika dikaji lebih lanjut berdasarkan analisis secara statistika, **garis tersebut memiliki kelemahan** karena hanya menjelaskan varians titik-titik yang kecil



Model Summary and Parameter Estimates

Dependent Variable: Y

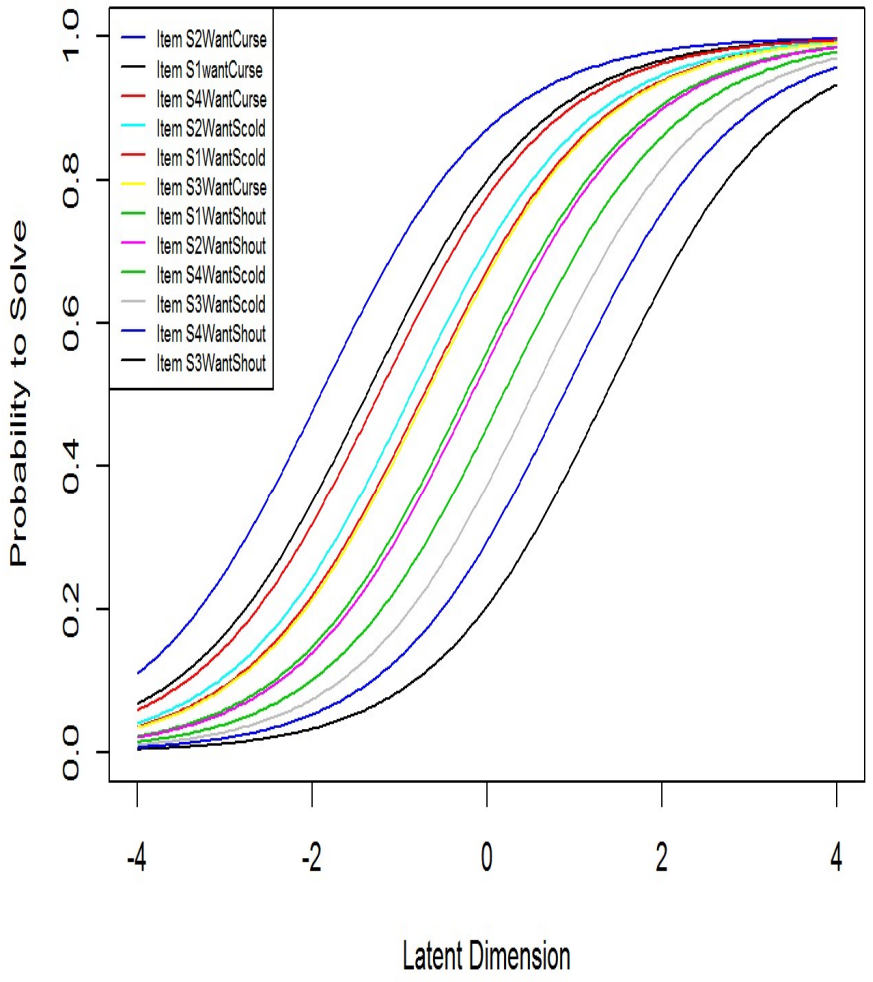
	Model Summary		
Equation	R Square	F	df1
Linear	.389	42.189	1
Quadratic	1.000	574072.500	2

The independent variable is X.

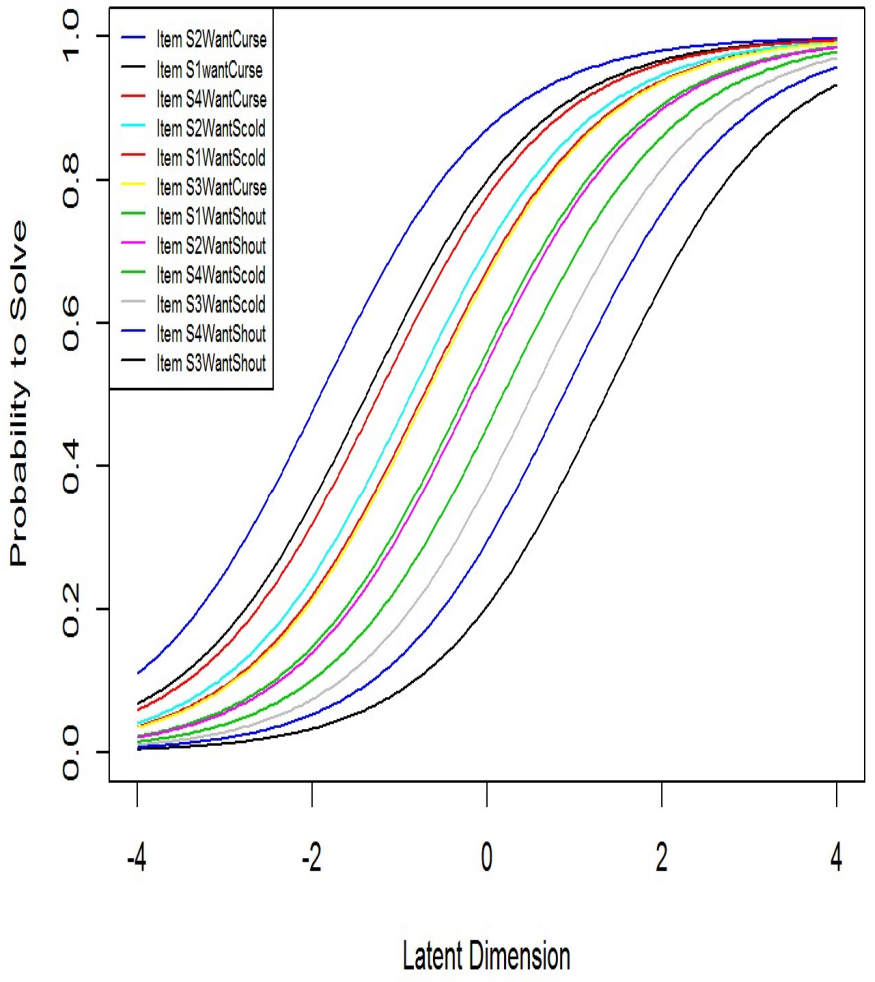
Model Linier menjelaskan 38,9 persen varians tetapi model kuadratik menyumbangkan 100 persen varians

Pelajaran yang diambil adalah kita tidak bisa melihat ketepatan model dengan melihat melalui garis yang dihasilkan dari kurva karakteristik butir.

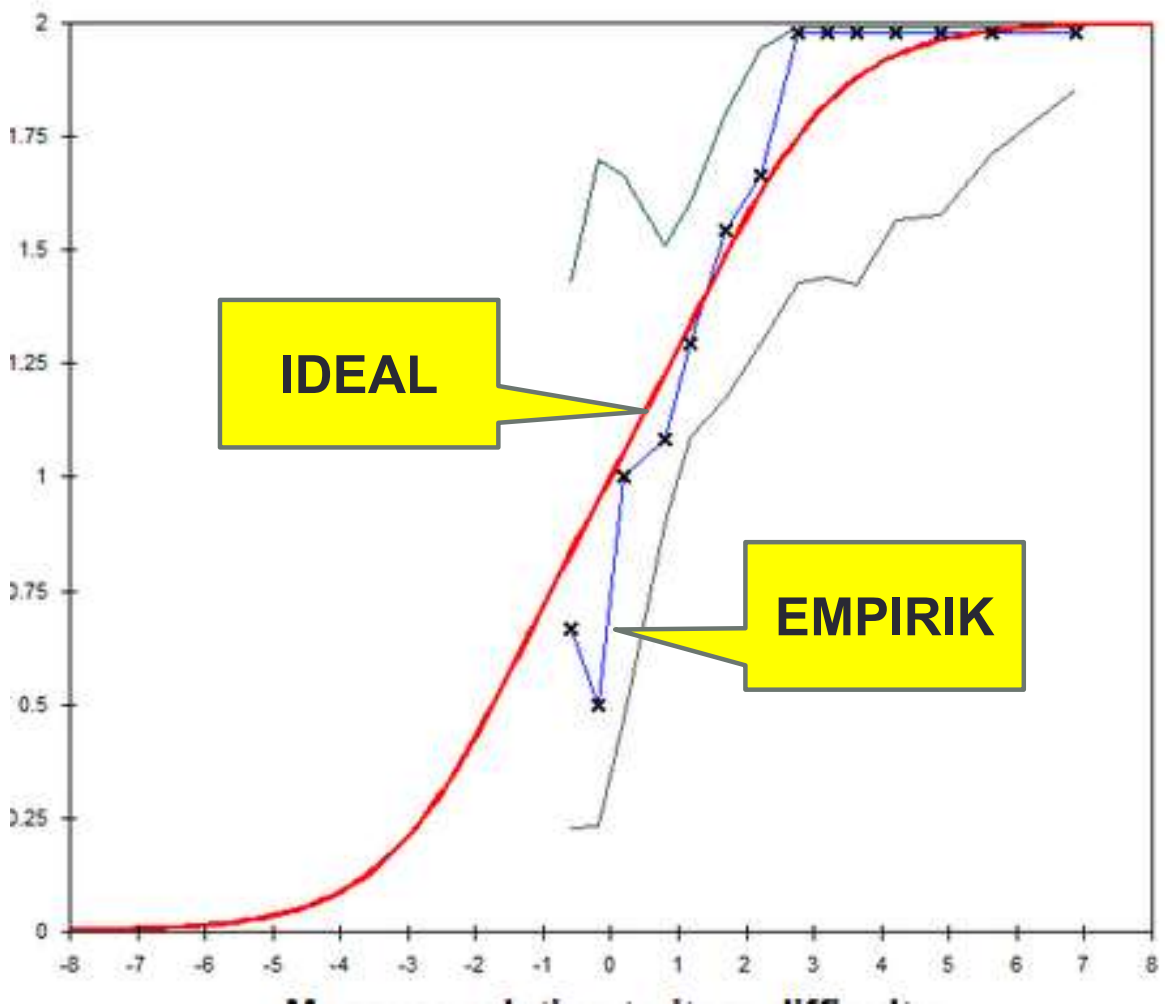
Ketepatan model harus dilihat melalui **Uji Statistik**. Gambar yang bagus dari sebuah kurva yang dihasilkan oleh Model Rasch belum tentu menunjukkan butir tersebut sesuai dengan model.



- Contoh disamping ini menunjukkan dari 12 butir dalam sebuah tes. Semua butir terlihat **IDEAL** karena mengikuti model Rasch.
- Tidak ada butir yang memiliki daya diskriminasi yang negatif atau rendah, karena Model Rasch tidak mengenal adanya variasi daya diskriminasi
- Bagi Model Rasch, daya diskriminasi bukan properti penting pada sebuah pengukuran.



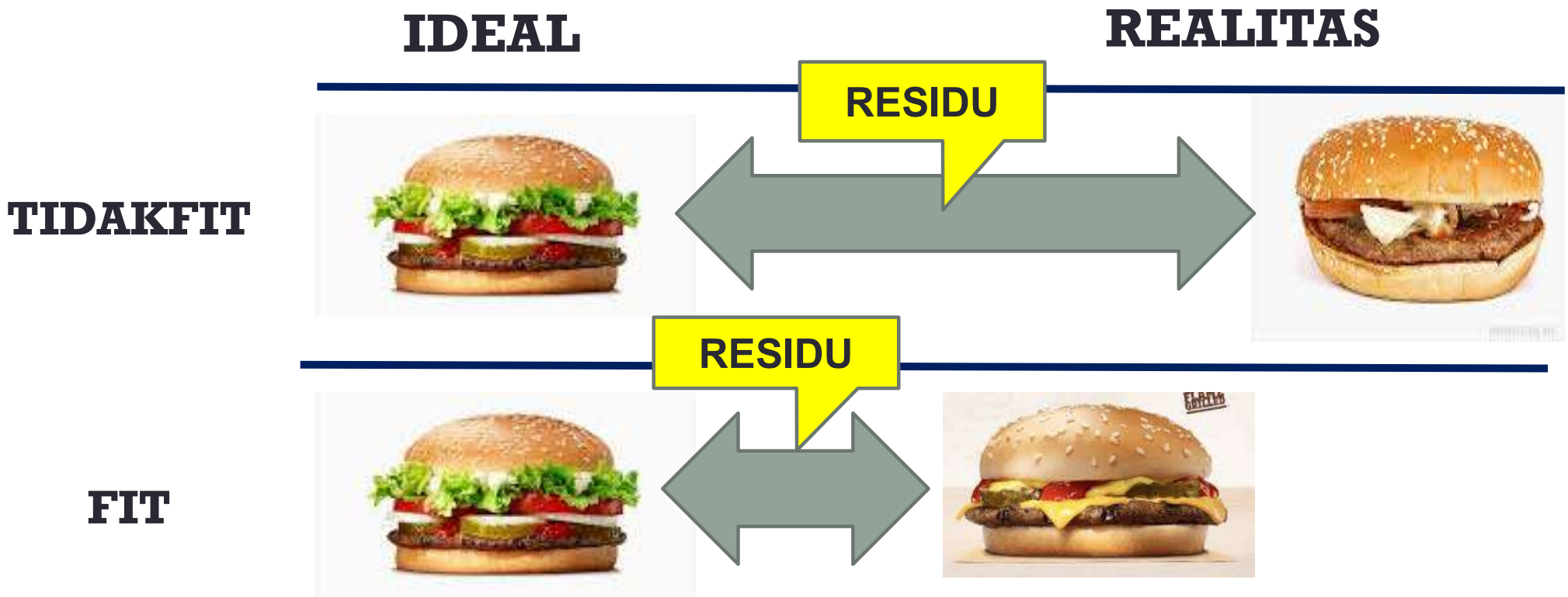
- Pengujian model fit melalui grafik karakteristik butir dapat dilakukan dengan membandingkan antara sesuatu yang jadi standar **(ideal)** dengan sesuatu yang didapatkan dari data atau fenomena (**empirik).**
- Grafik disamping ini menunjukkan perbandingan tersebut. **Garis Berwarna Merah** menunjukkan sesuatu yang ideal dan **Garis Berwarna Biru** menunjukkan sesuatu yang empirik (berdasarkan data di lapangan)



- Sebuah butir dapat dikatakan KURANG FIT ketika jauh dari sesuatu yang ideal dan dapat dikatakan FIT jika mendekati kondisi IDEAL.

	IDEAL	REALITAS
TIDAK FIT		
FIT		

- Dalam Analisis berbasis Rasch dikenal dengan residu yang sesuatu yang acak yang tidak bisa dijelaskan oleh model Rasch. Semakin jauh sebuah butir dari kondisi ideal, semakin banyak residunya.



Berdasarkan Model Rasch, butir yang memiliki **PERFORMANSI IDEAL** atau butir yang lolos dalam seleksi butir adalah butir yang **FIT dengan MODEL RASCH**

APA KOEFISIEN YANG DIPAKAI?

Kai Kuadrat

1

- Kai Kuadrat menguji perbedaan antara kondisi IDEAL dan EMPIRIK.
- Hasil uji kai kuadrat yang signifikan ($p < 0,05$) menunjukkan ada perbedaan yang signifikan antara kondisi ideal dan empirik.
- Artinya, butir yang diperiksa memiliki performansi yang kurang optimal

MicroCAT (tm) Testing System

Copyright (c) 1982 - 1995 by Assessment Systems Corporation

Rasch Model Item Calibration Program -- RASCAL for
Date: 06/06/95

Final Parameter Estimates for Data from File C:\RAS

Cut off dari harga Kai kuadrat dengan df = 10 adalah 18,30 (lihat tabel)

Item	a	b	c	SE(b)	Chi Sq.	df	Scaled Diff
1	0.657	-1.319	0.000	0.138	14.727	10	88
2	0.657	-0.669	0.000	0.124	13.004	10	94
3	0.657	-2.884	0.000	0.229	25.404	10	74
4	0.657	-2.355	0.000	0.186	8.778	10	79
5	0.657	-3.575	0.000	0.313	29.069	10	67
6	0.657	-0.752	0.000	0.125	11.250	10	93
7	0.657	-2.709	0.000	0.213	48.363	10	75
8	0.657	-1.751	0.000	0.153	17.837	10	84
9	0.657	-1.751	0.000	0.153	14.085	10	84
10	0.657	-0.466	0.000	0.122	14.841	10	96

- Uji ketepatan model melalui kuadrat memiliki keterbatasan karena sangat tergantung dari jumlah ukuran sampel.

- Semakin besar ukuran sampel uji kai kuadrat lebih mudah mendapatkan hasil yang signifikan, sebagaimana halnya uji statistika yang lain.

- Oleh karena itu pendekatan ini jarang dipakai. Uji kai kuadrat hanya dipakai sebagai pemeriksaan biasa yang tidak menentukan penguguran butir dalam analisis butir.

Infit & Outfit

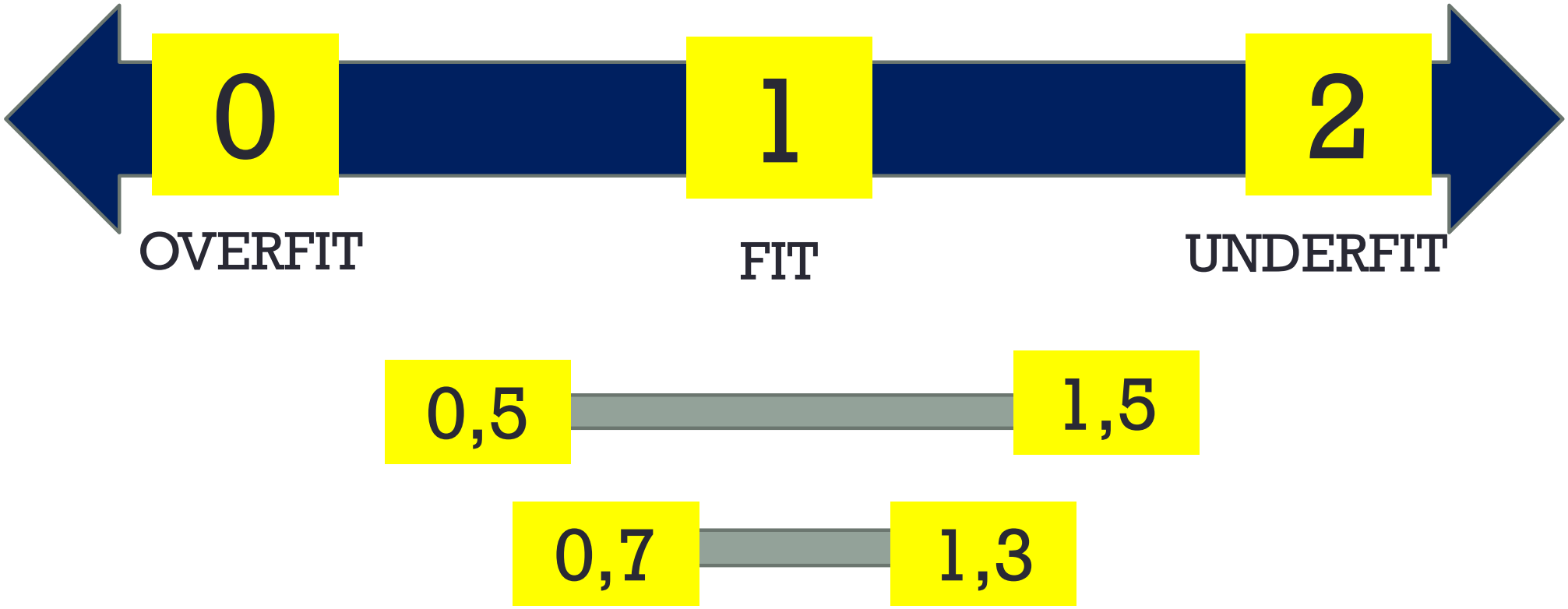
2

Infit dan outfit merupakan koefisien yang populer dalam menunjukkan ketepatan sebuah butir dengan Model Rasch.

Koefisien ini didapatkan dari **RESIDU** (gap antara IDEAL – DATA) yang dirata-ratakan sehingga semakin kecil menunjukkan sedikitnya residu

OVERFIT

Harga INFIT dan OUTFIT bergerak antara dua harga terhingga namun seringkali bergerak antara 0 hingga 2 dengan titik kritis adalah 1. Semakin rendah menunjukkan sebuah butir semakin fit dengan model dan sebaliknya



Harga yang direkomendasikan dalam mengidentifikasi butir yang sesuai dengan model bermacam-macam. ada para ahli yang mengatakan bahwa nilai yang direkomendasikan adalah 0,5 sampai 1,5, ada juga yang merekomendasikan antara 7 sampai 1,3 tergantung dari nilai kepentingan pengukuran yang dilakukan. Pengukuran yang dilakukan untuk kepentingan seleksi menetapkan interval harga infit atau outfit yang lebih

INFIT vs OUTFIT

- Harga koefisien OUTFIT sangat dipengaruhi oleh adanya outliers (orang yang pintar tidak bisa menjawab soal yang mudah dan sebaliknya).
- Nilai INFIT cenderung tidak terpengaruh terhadap outliers namun terpengaruh oleh adanya gap antara kemampuan orang dengan butir yang memiliki tingkat kesulitan sama dengan kemampuan orang tersebut. Gap inilah yang memang merupakan informasi yang diharapkan.

Saran dalam Mengevaluasi Ketepatan Butir (Linacre, 2011)

1. Investigate **negative point-measure** or **point-biserial correlations**

   Korelasi biserial yang negatif dapat menunjukkan bahwa adanya salah penyekoran karena salah kunci atau salah kode. Indeks infit atau outfit kurang mampu mendeteksi masalah ini

Saran dalam Mengevaluasi Ketepatan Butir (Linacre, 2011)

2. Investigate **outfit** before **infit**

   Meskipun informasi utama adalah koefisien infit akan tetapi investigasi melalui outfit juga diperlukan untuk menunjukkan apakah butir yang dikaji menghasilkan banyak outliers yang bisa jadi muncul karena salah konsep, tingkat kesulitan yang terlalu mudah atau terlalu sulit.

Saran dalam Mengevaluasi Ketepatan Butir (Linacre, 2011)

3. mean-square before t-standardized

Nilai koefisien infit-outfit memiliki dua jenis, yaitu nilai mentah (MNSQ) dan harga terstandarisasi (Z-MNSQ). Harga terstandarisasi sangat terpengaruh oleh ukuran sampel dsb sehingga seringkali menganggap butir sebagai butir yang jelek (Z-MNSQ > 3) padahal sebenarnya baik. Sebagian besar peneliti langsung melihat harga INFIT dan OUTFIT daripada harga terstandarisasinya

Saran dalam Mengevaluasi Ketepatan Butir (Linacre, 2011)

3. high values before low or negative values

Harga infit – outfit yang rendah menunjukkan butir tersebut tumpang tindih dengan butir lain sedangkan yang harganya tinggi menunjukkan banyak faktor acak yang mempengaruhi jawaban individu terhadap butir tersebut.

Saran dalam Mengevaluasi Ketepatan Butir

1. Investigate **negative point-measure** or **point-biserial correlations**
2. Investigate **outfit** before **infit**
3. **mean-square** before **t-standardized**
4. **high values** before **low or negative values**.

3

Q-Index

Indeks Q dikembangkan oleh Rost and von Davier pada tahun 1994. Indeks ini dikembangkan untuk melengkapi indeks **infit** dan **outfit**.

Nilai indeks Q berada antara 0 hingga 1. Nilai di bawah 0,5 menunjukkan butir memiliki fit dengan model Rasch sedangkan di atas 0,5 menunjukkan sebaliknya

CONTOH PENULISAN LAPORAN

Heinz, dkk. (2009). Caffeine Expectancy. Instrument Development in the Rasch Measurement Framework. **Psychology of Addictive Behaviors**. Vol. 23, No. 3, 500 – 511

- Linacre (2004) suggested that **a standardized outfit mean-square statistic** for both persons and items in **excess of 2.00** indicates the presence of excessive noise in the observations. For misfit items, the number of large standardized residuals contributing to the overall misfit was counted for each of the positively misfit items

Neu dkk. (2010)

Our results show good outfit mean square statistics for all items of the ESS, ranging from 0.62 to 1.22. Infit mean squares range from 0.77 to 1.21 (table 2). Moreover, all observed values are less than 1.4, which indicates the unidimensional nature of the sleepiness measure from the ESS.

Table 2. Rasch item statistics of the ESS

Item content	Logits	SE	Infit mean squares	Outfit mean squares
(5) Lying down to rest in the afternoon when circumstances permit	−1.49	0.12	0.95	1.02
(2) Watching TV	−0.51	0.11	1.12	1.19
(4) As a passenger in a car for 1 h without a break	0.14	0.11	1.21	1.22
(1) Sitting and reading	0.39	0.11	1.00	0.96
(7) Sitting quietly after lunch without alcohol	1.04	0.12	0.77	0.77
(3) Sitting inactively in a public place	1.32	0.12	0.90	0.98
(8) In a car, while stopped for a few minutes in traffic	3.33	0.20	0.89	0.62
(6) Sitting and talking to someone	3.94	0.24	0.95	1.00

Cheng dkk. (2011)

TABLE 2. Item Fit Statistics

Question no.	Items	Logit measure	Item fit statistics of the original 31-item: INFIT MNSQ	Item fit statistics of the original 31-item: OUTFIT MNSQ
Symptom				
OMQoL4	Oozing/bleeding on lips, or inside mouth.	2.10	**1.56**	**2.61**
OMQoL1	Swelling inside mouth	−0.14	1.07	1.12
OMQoL2	Mouth ulcer	−0.38	1.01	1.04
OMQoL9	Difficulty in opening mouth	0.91	1.21	1.05
OMQoL5	Discomfort while tooth brushing/mouth rinsing	−0.36	1.05	1.04
OMQoL8	Redness or burning sensation inside mouth	−0.73	0.76	0.81
OMQoL7	Mouth pain	−0.61	0.78	0.76
OMQoL6	Trouble sleep associated with mouth pain	−0.17	1.16	1.08
OMQoL3	Distress associated with mouth pain	−0.62	0.79	0.78

CONTOH ANALISIS

http://langtest.jp/shiny/rasch/

1. Siapkan data di EXCEL yang sudah di skor 0 dan 1 (Benar Salah atau YA TIDAK)
2. Copy Paste ke dalam Kolom yang disediakan oleh link di atas

RASCH MODEL

Bagian 3

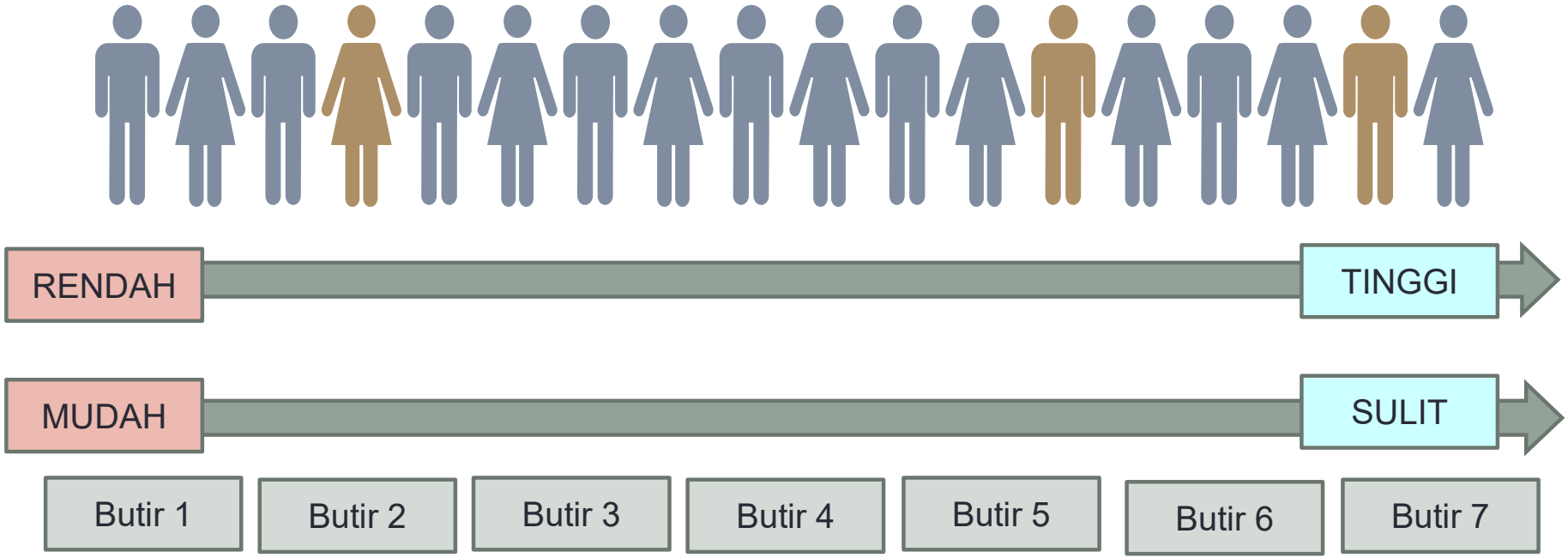
Peta Butir Orang

Agar pemirsa video ini dapat memahami dengan lebih lengkap, syaratnya adalah telah memahami konsep mengenai eror standar atau eror pengukuran.

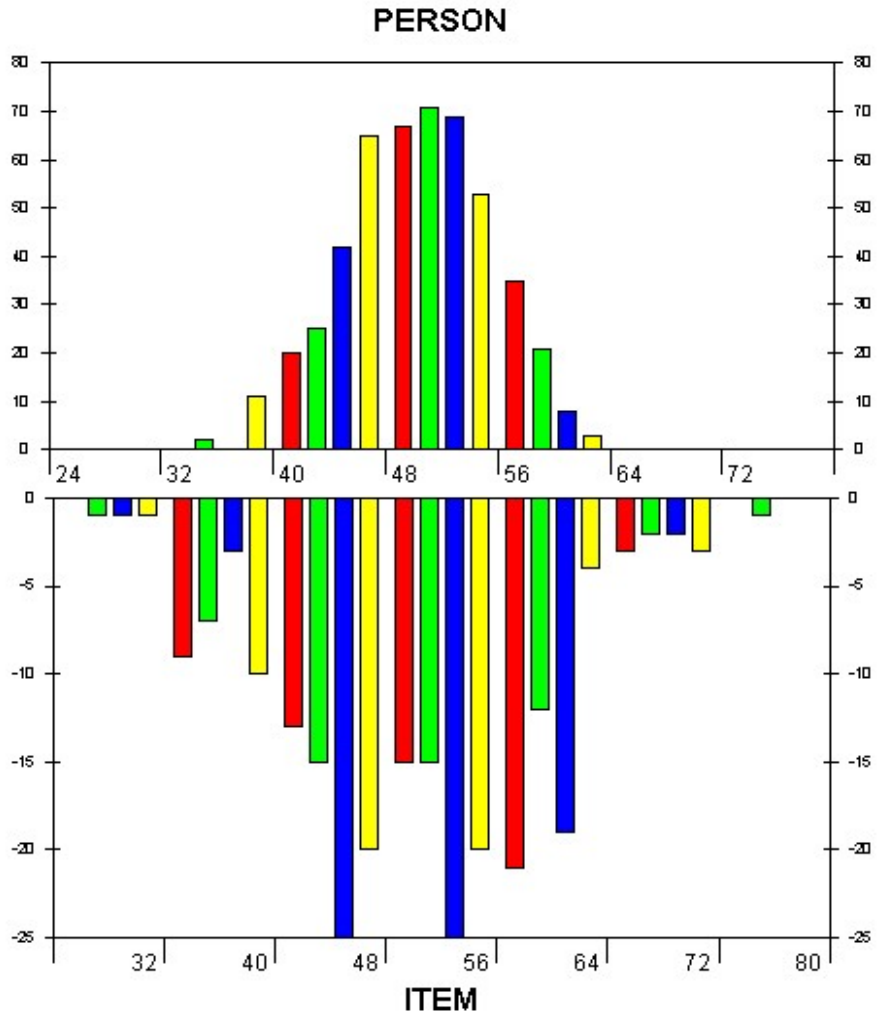
Silahkan Anda melihat video dengan judul Deviasi Standar dan Eror Standar pada Channel kami

Peta Orang-Butir
(Item Person Map)

Pada video sebelumnya kita telah belajar bahwa parameter orang dan butir dalam Rasch Model telah dijadikan ‘satukan’ sehingga keduanya memiliki satuan yang sama dalam bentuk LOGIT



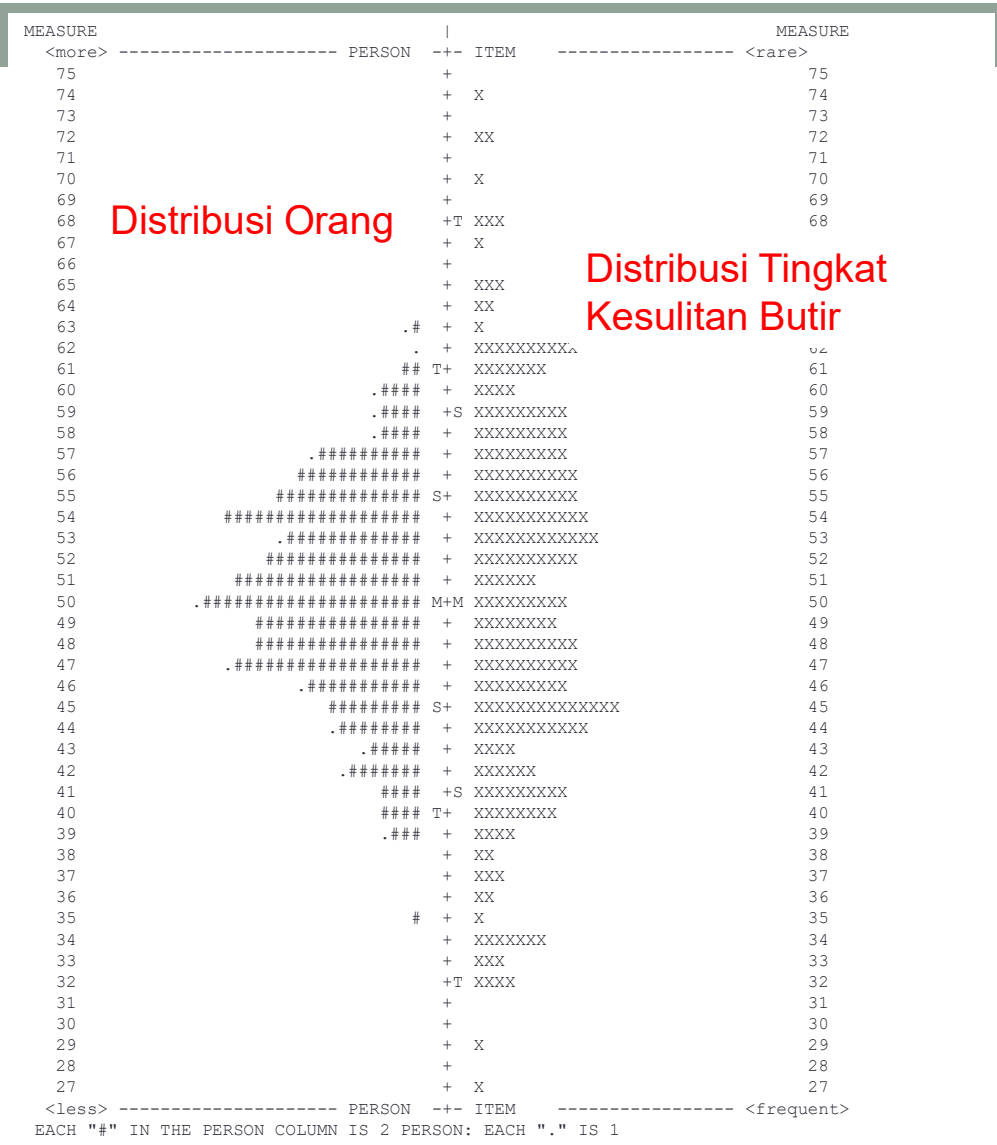
Jika keduanya disatukan, jadilah PETA BUTIR ORANG atau yang lebih dikenal denga WRIGHT MAP karena yang mengembangkan adalah Benjamin Wright



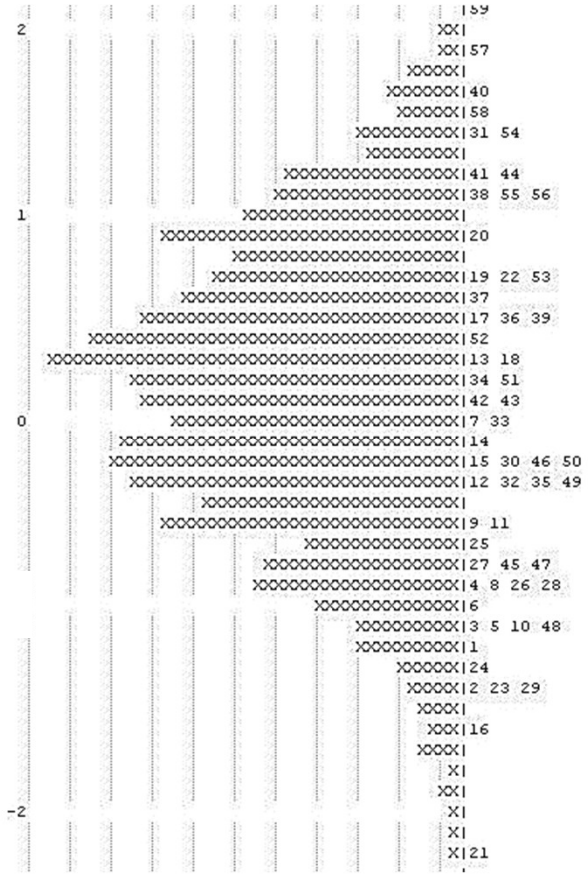
IDEALNYA ....

Idealnya butir-butir memiliki sebaran tingkat kesulitan yang bervariasi dan menjangkau sebaran individu seperti pada gambar di samping ini. Gampangannya, dalam satu tingkat kemampuan orang harus ada butir yang menempati pada titik tersebut.

Pada gambar di samping ini satu buah kode # mewakili 5 orang dan satu buah x mewakili 1 butir.



Kemampuan Tinggi



Orang

Kemampuan Rendah

Tingkat Kesulitan Tinggi

Butir

Tingkat Kesulitan Rendah

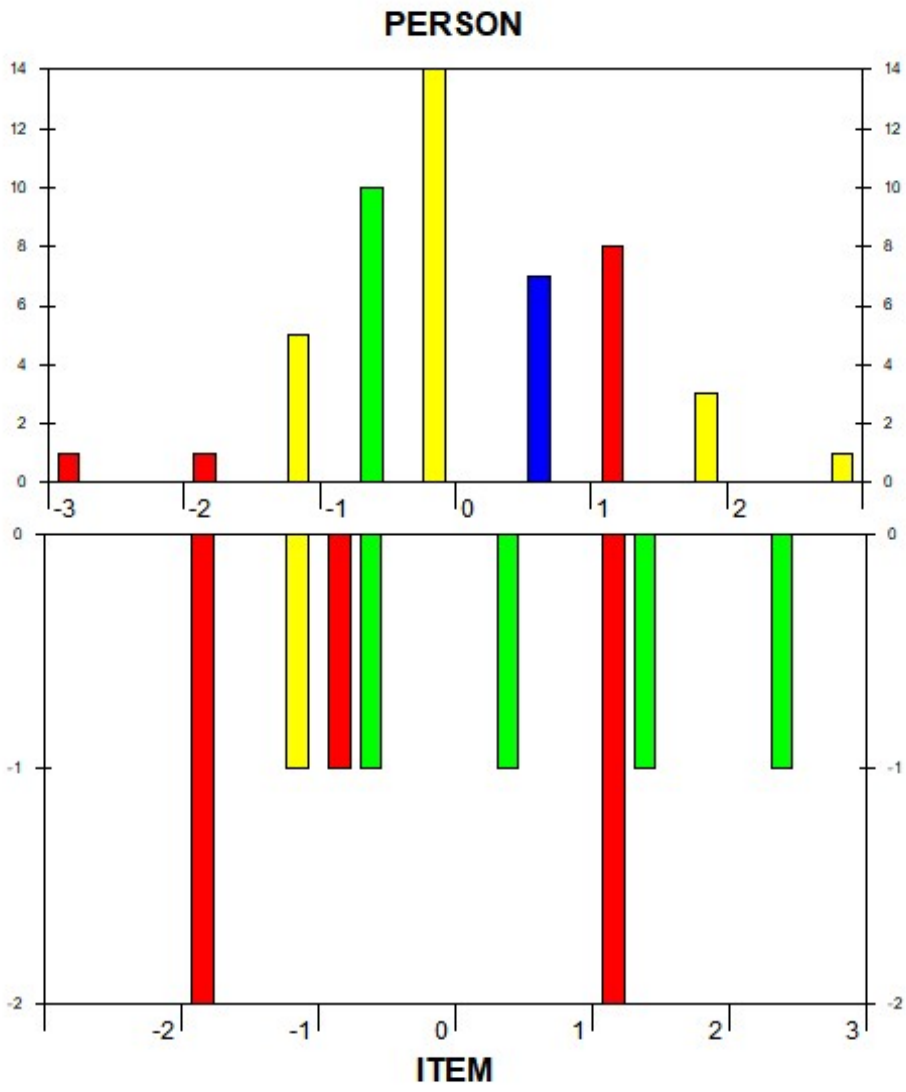
RASCH MODEL

Bagian 4

Indeks Separabilitas dan Reliabilita

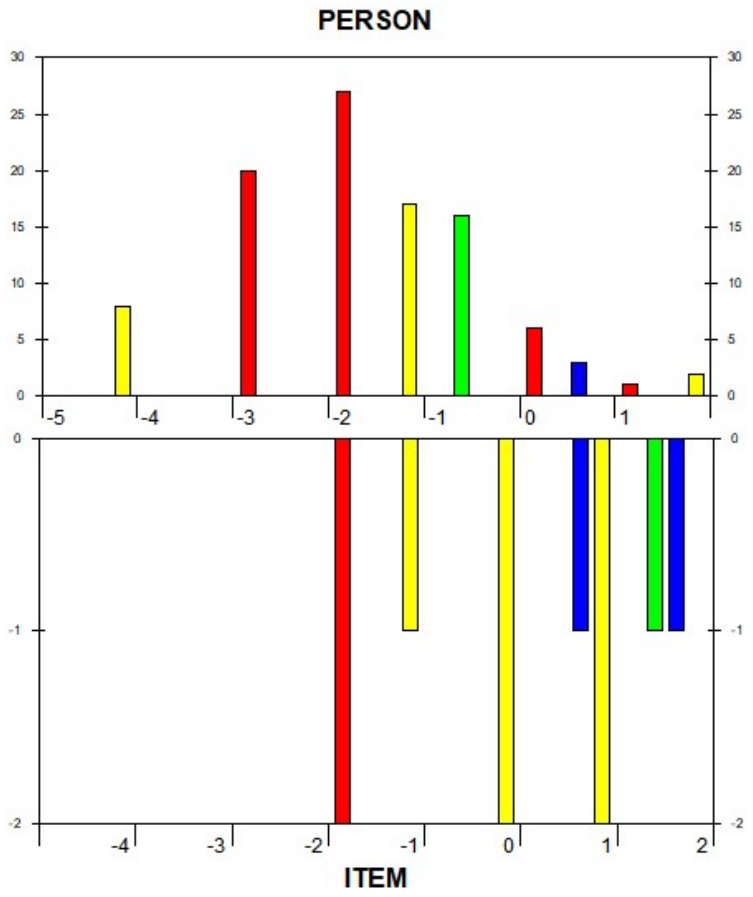
PERSON	50 INPUT		50 MEASURED		INFIT		OUTFIT	
	TOTAL	COUNT	MEASURE	REALSE	IMNSQ	ZSTD	OMNSQ	ZSTD
MEAN	5.2	10.0	.10	.86	.98	.0	1.03	.1
S.D.	1.6	.0	1.06	.10	.38	1.0	.68	1.0
REAL RMSE	.86	TRUE SD	.61	SEPARATION	.71	PERSON RELIABILITY		.33

ITEM	10 INPUT		10 MEASURED		INFIT		OUTFIT	
	TOTAL	COUNT	MEASURE	REALSE	IMNSQ	ZSTD	OMNSQ	ZSTD
MEAN	25.9	50.0	.00	.37	1.00	-.1	1.03	.1
S.D.	12.6	.0	1.42	.06	.11	.5	.18	.5
REAL RMSE	.38	TRUE SD	1.37	SEPARATION	3.64	ITEM RELIABILITY		.93



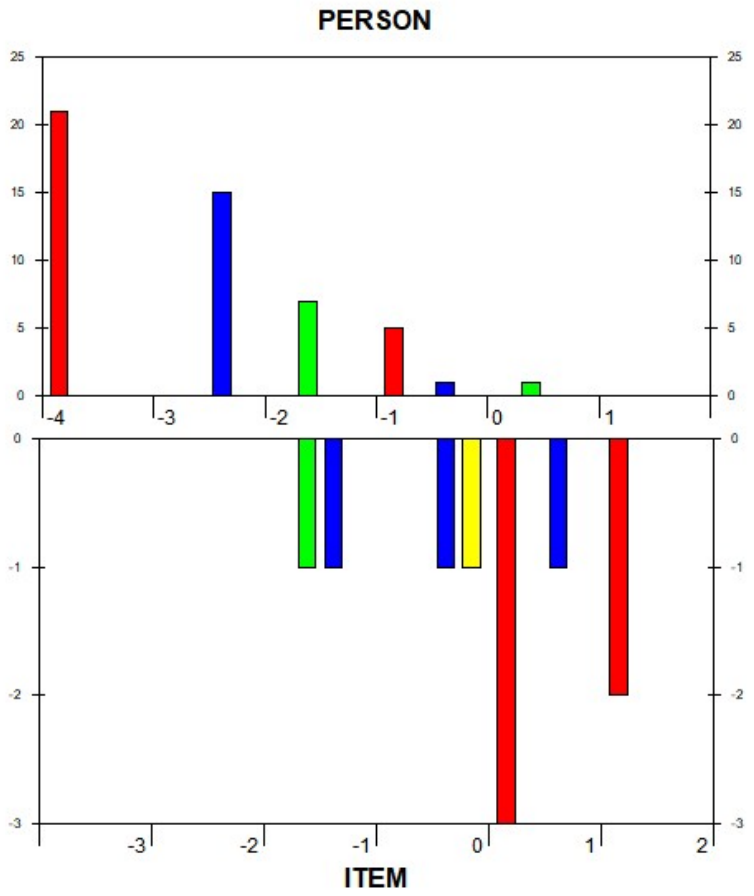
PERSON	100 INPUT		100 MEASURED		INFIT		OUTFIT	
	TOTAL	COUNT	MEASURE	REALSE	IMNSQ	ZSTD	OMNSQ	ZSTD
MEAN	2.6	10.0	-1.58	1.02	.99	.0	.94	.1
S.D.	1.7	.0	1.28	.30	.39	.9	.74	.9
REAL RMSE	1.06	TRUE SD	.73	SEPARATION	.69	PERSON RELIABILITY		.32

ITEM	10 INPUT		10 MEASURED		INFIT		OUTFIT	
	TOTAL	COUNT	MEASURE	REALSE	IMNSQ	ZSTD	OMNSQ	ZSTD
MEAN	26.0	100.0	.00	.31	1.01	.0	.94	-.2
S.D.	17.6	.0	1.23	.07	.16	1.2	.22	1.0
REAL RMSE	.32	TRUE SD	1.19	SEPARATION	3.73	ITEM RELIABILITY		.93



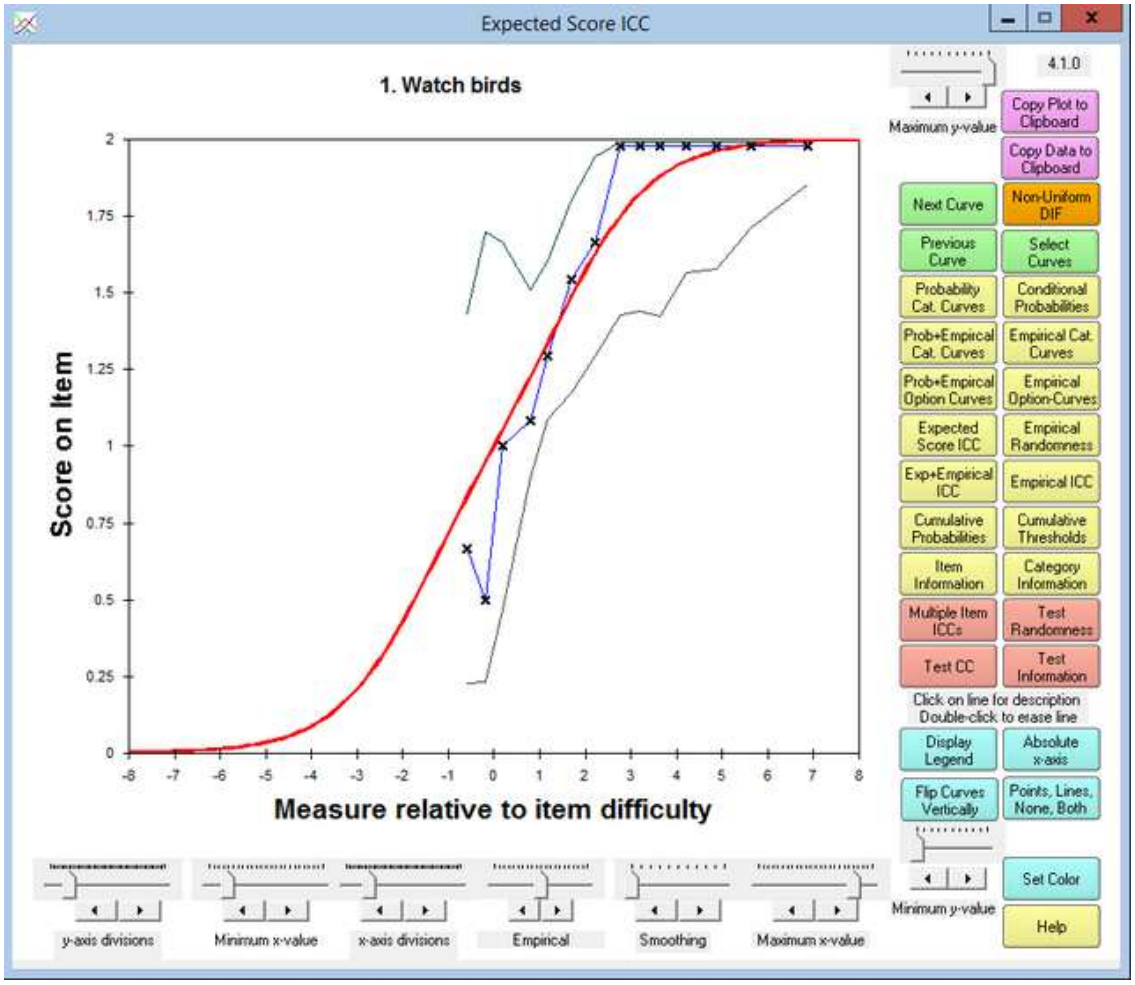
PERSON	50 INPUT		50 MEASURED		INFIT		OUTFIT	
	TOTAL	COUNT	MEASURE	REALSE	IMNSQ	ZSTD	OMNSQ	ZSTD
MEAN	1.1	10.0	-2.67	1.36	.99	.2	.92	.2
S.D.	1.3	.0	1.15	.45	.29	.7	.67	.7
REAL RMSE	1.43	TRUE SD	.00	SEPARATION	.00	PERSON RELIABILITY		.00

ITEM	10 INPUT		10 MEASURED		INFIT		OUTFIT	
	TOTAL	COUNT	MEASURE	REALSE	IMNSQ	ZSTD	OMNSQ	ZSTD
MEAN	5.4	50.0	.00	.58	.99	.2	.92	.1
S.D.	3.5	.0	.85	.12	.16	.6	.37	.9
REAL RMSE	.59	TRUE SD	.60	SEPARATION	1.02	ITEM	RELIABILITY	.51



Pelajaran yang diambil adalah kita tidak bisa melihat ketepatan model dengan melihat melalui garis yang dihasilkan dari kurva karakteristik butir.

Ketepatan model harus dilihat melalui **Uji Statistik**. Gambar yang bagus dari sebuah kurva yang dihasilkan oleh Model Rasch belum tentu menunjukkan butir tersebut sesuai dengan model.



Sekilas Teori PSIKOMETRI

- **Teori Tes Psikometri**
  - Teori Tes Klasik
  - Teori Tes Modern (Teori Respons Butir)
    - Rasch Model

- **Statistika**
  - Uji hipotesis bivariat
  - Uji hipotesis multivariat (model)

Perbandingan Tes

Norm Referenced

1. Menggunakan norma untuk memaknai skor
2. Tinggi rendahnya skor berdasarkan posisi relatif dari orang lain
3. Menekankan pada perbedaan orang (*individual differences*)

Criterion Referenced

1. Menggunakan kriteria untuk memaknai skor
2. Tinggi rendahnya skor didasarkan pada (level) capaian.
3. Menekankan pada pembuktian performansi yang mendukung suatu tugas

Table 3. Comparative Ideals for NR and CR Interpretive Situations

	Norm-Referenced	Criterion-Referenced
Item difficulty	Moderate	Easy-to-hard
Item discrimination	High positive	Nonnegative
Score variability	Maximize	Non-issue
Error estimate	High reliability coefficient	High decision consistency index

Pengantar

- **Pendekatan Thurstone**
  - Penskalaan dilakukan pada stimulus
  - Stimulus memiliki bobot
- **Pendekatan Guttman**
  - Penskalaan dilakukan pada stimulus
  - Stimulus memiliki tidak memiliki bobot
- **Pendekatan Likert**
  - Penskalaan dilakukan pada respons
  - Menekankan pada akumulasi stimulus

Skor Butir – Perbandingan Penskalaan

Format Skala	Indikator	Penyekoran
Likert (Summated Ratings)	Tidak terbobot	Tidak terbobot
Guttman	Terbobot	Tidak terbobot
Thurstone	Terbobot	Terbobot

Item	Likert	Gutt	Thurs
1. Memelototi	5	1	1.2
2. Memukul	5	1	3
3. Menendang	5	1	4.2
4. Mencederai	5	1	5.5

- Skala Likert → IRT → Penyekoran Terbobot
- Dengan menggunakan pemodelan IRT, skor yang didapatkan individu telah terbobot betdasarkan parameter butir
- Item yang memiliki daya diskriminasi tinggi mendapatkan bobot yang tinggi

FOCAL ARTICLE

75 Years After Likert: Thurstone Was Right!

FRITZ DRASGOW
University of Illinois at Urbana-Champaign

OLEKSANDR S. CHERNYSHENKO
Nanyang Technological University

STEPHEN STARK
University of South Florida

Abstract

For over three-quarters of a century researchers and practitioners have analyzed rating scale da that assume a dominance response process wherein an individual high on the trait assessed is a positively with high probability. This approach derives from Likert's famous 1932 approach to and analysis of rating scales. In this paper, we argue that Likert scaling and related metho Instead, we propose that methods that have evolved from Thurstone (1927, 1928, 1929) scalin representation of the choice process underlying rating scale judgments. These methods hyp point response process where the probability of endorsement is assumed to be directly related t the statement to the individual's standing on the assessed trait. We review some research showi of ideal point methods for personality assessment and then describe several settings in which id should provide tangible improvements over traditional approaches to assessment.

Industrial and Organizational Psychology, **3** (2010), 481–484.
Copyright © 2010 Society for Industrial and Organizational Psychology. 1754-9426/10

Developing and Scaling Personality Measures: Thurstone Was Right—But So Far, Likert Was Not Wrong

FREDERICK L. OSWALD
Rice University

KRAIG L. SCHELL
Angelo State University

Although we are in agreement with Drasgow, Chernyshenko, and Stark (2010) on the appropriateness of considering ideal point models in psychological testing, we worded at the extremes like this, both the dominance response model and the ideal point model yield essentially the same personality trait scores, as Drasgow, et al.

Teori Skor Murni Klasik

- **Deskripsi**
  - Teori yang menjelaskan skor tes yang didalamnya memuat skor murni dan eror pengukuran.

- **Asumsi**

- a common-interval linear scale
- The property of specific objectivity
- sample distribution–free estimation

- A key concern in psychometrics is whether scores are consistent (i.e., score reproducibility across different test situations). This has led to the development of several CTT indices of reliability depending on test administration design.

Perbandingan Tes Tradisional dan Modern

Model Tradisional

1. Eror standar pengukuran (SEM) berlaku untuk populasi.
2. Tes yang lebih panjang lebih reliabel dibanding yang pendek
3. Membandingkan skor dapat dilakukan jika menggunakan tes yang paralel.
4. Parameter butir tergantung dari kerepresentatifan sampel
5. Skor tes dimaknai dengan cara membandingkan posisinya di grup

Model Modern

1. SEM berbeda berdasarkan pola responsnya tapi dapat digeneralisasikan pada populasi
2. Tes dengan ukuran pendek dapat lebih reliabel dibandingkan dengan yang panjang
3. Membandingkan skor tes yang berasal dari format yang berbeda akan optimal jika tingkat kesulitannya bervariasi
4. Parameter butir yang tidak bias dapat dihasilkan dari sampel yang tidak representatif
5. Skor tes memiliki makna ketika mereka dibandingkan dari jaraknya dengan butir

Model Tradisional

6. Skala interval dapat dicapai ketika skor terdistribusi normal
7. Butir dengan format campuran menyebabkan ketidakseimbangan terhadap skor total
8. Perubahan skor (*change score*) tidak dapat dibandingkan ketika skor awalnya berbeda
9. Faktor analisis pada butir dikotomi menghasilkan artefak daripada faktor
10. Fitur stimulus pada butir tidak penting untuk dibandingkan

Model Modern

- Interval scale properties are achieved by applying justifiable measurement models.
- Mixed item formats can yield optimal test scores.
- Change scores can be meaningfully compared when initial score levels differ.
- Factor analysis on raw item data yields a full information factor analysis.
- Item stimulus features can be directly related to psychometric properties.

PROGRAM ITEMAN

PROGRAM ITEMAN

Kaidah Penulisan Kode Iteman

- **Kolom 1-3** merupakan respon atau jumlah aitem yang akan dianalisis
- **Kolom 4**, blank, dibiarkan kosong (hanya di spasi)
- **Kolom 5**, kode OMIT, yaitu apabila ada subjek yang tidak menjawab pada suatu aitem mau di skor apa, biasanya ditulis dengan huruf O (bukan nol)
- **Kolom 6**, blank, dibiarkan kosong (hanya di spasi)
- **Kolom 7**, adalah informasi untuk soal yang tidak dapat diselesaikan subjek karena waktu habis (NOT REACH), ditulis N
- **Kolom 8**, blank. Dibiarkan kosong (hanya di spasi)
- **Kolom 9-10** adalah jumlah data awal yang tidak akan diikutsertakan dalam analisis. Apabila semua data akan dianalisis maka pada baris tersebut diisi 0 (nol).

Contoh Kode Iteman

Kunci

```
 10 O N 10
EDCDADCCDD -> kunci
5555555555 -> jumlah opsi
YYYYYYYYYY -> Y artinya di analisis

Andhika    ECEDAEAEEEC
Afif    BAODAOADDCA
Agnindia    BABBCDDDCBD
Aldya   BAAAAAADACE
```

DIMENSIONALITAS

Tahap Tahap

- **Membuat Kode**
  - Buka Winsteps
  - Pilih SELECT FILE, lalu pilih File EXCEL/SPSS.
  - Pilih variabel identitas, lalu masukkan ke **Person Label Variables.**
  - Pilih variabel butir, lalu masukkan ke **Item Response Variables.**
  - Klik CONSTRUCT WINSTEPS FILE, beri nama filenya lalu klik SAVE
  - Dibuka file tersebut.
    - Di bawah NAMELEN, tulis KEY=11111111111 dst (sejumlah butir soal atau lebih). Lalu SAVE
  - Pilih file yang terakhir. Klik kanan lalu pilih OPEN WITH lalu pilih WINSTEPS.EXE. Lalu ENTER 2 kali.

Istilah

- **Tingkat Kesulitan dan**
  - Item Measure = Tingkat Kesulitan Butir
  - Person Measure= Kemampuan Orang
- **Indeks Ketepatan**
  - Infit (untuk butir dan person)
  - Outfit (untuk butir dan person)
  - Overfit, Fit, Misfit
- **Peta Orang Butir (Item Person Map)**

OUTPUT WINSTEP : Independensi Lokal

TABLE 23.99 (Klik Output Table 23 Dimensionality)

CORREL-ATION	ENTRY NUMBER	ITE	ENTRY NUMBER	ITE
.20	12	A12	13	A13
-.25	3	A3	11	A11
-.24	18	A18	19	A19
-.23	4	A4	12	A12
-.22	3	A3	5	A5
-.22	11	A11	15	A15
-.20	3	A3	9	A9
-.20	4	A4	25	A25
-.20	2	A2	16	A16
-.20	11	A11	18	A18

- Tabel ini menunjukkan 10 korelasi antar residu tertinggi.
- Korelasi yang tinggi menunjukkan bahwa butir yang bersangkutan berbagi residu yang cukup tinggi (ingat korelasi antar eror)
- Di tabel, tidak ada korelasi antar residu butir yang tinggi sehingga tidak ada kasus dependensi lokal

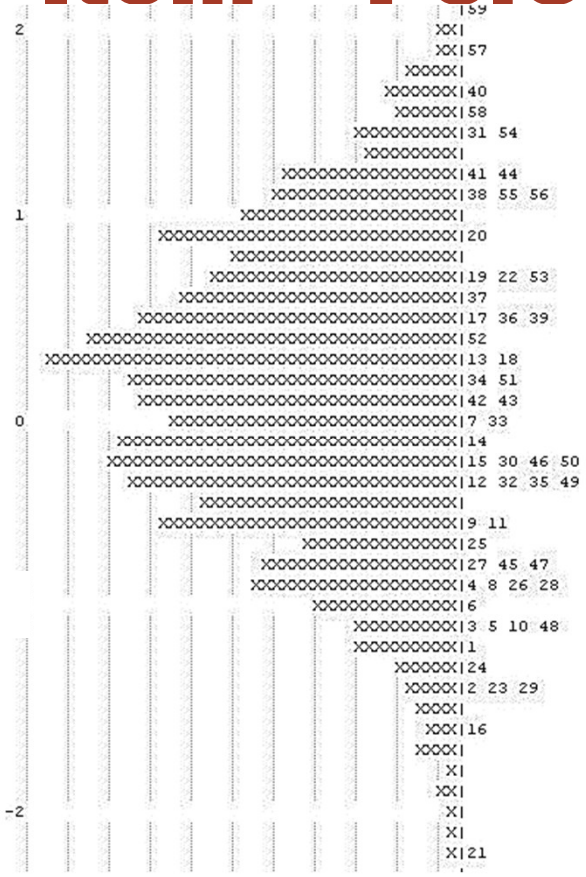
ITEM PERSON FIT

Item – Person Fit

Kemampuan Tinggi

Orang

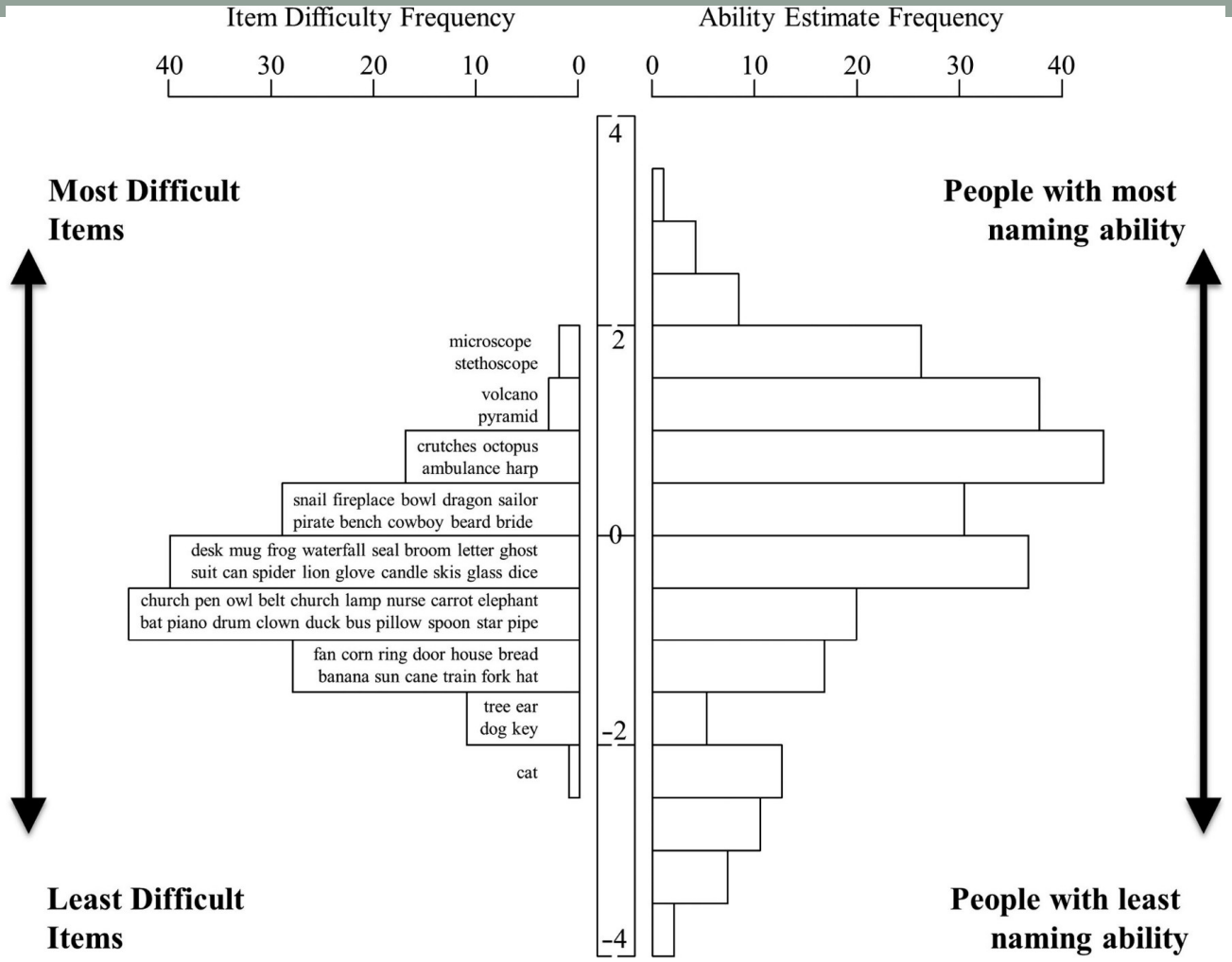
Kemampuan Rendah



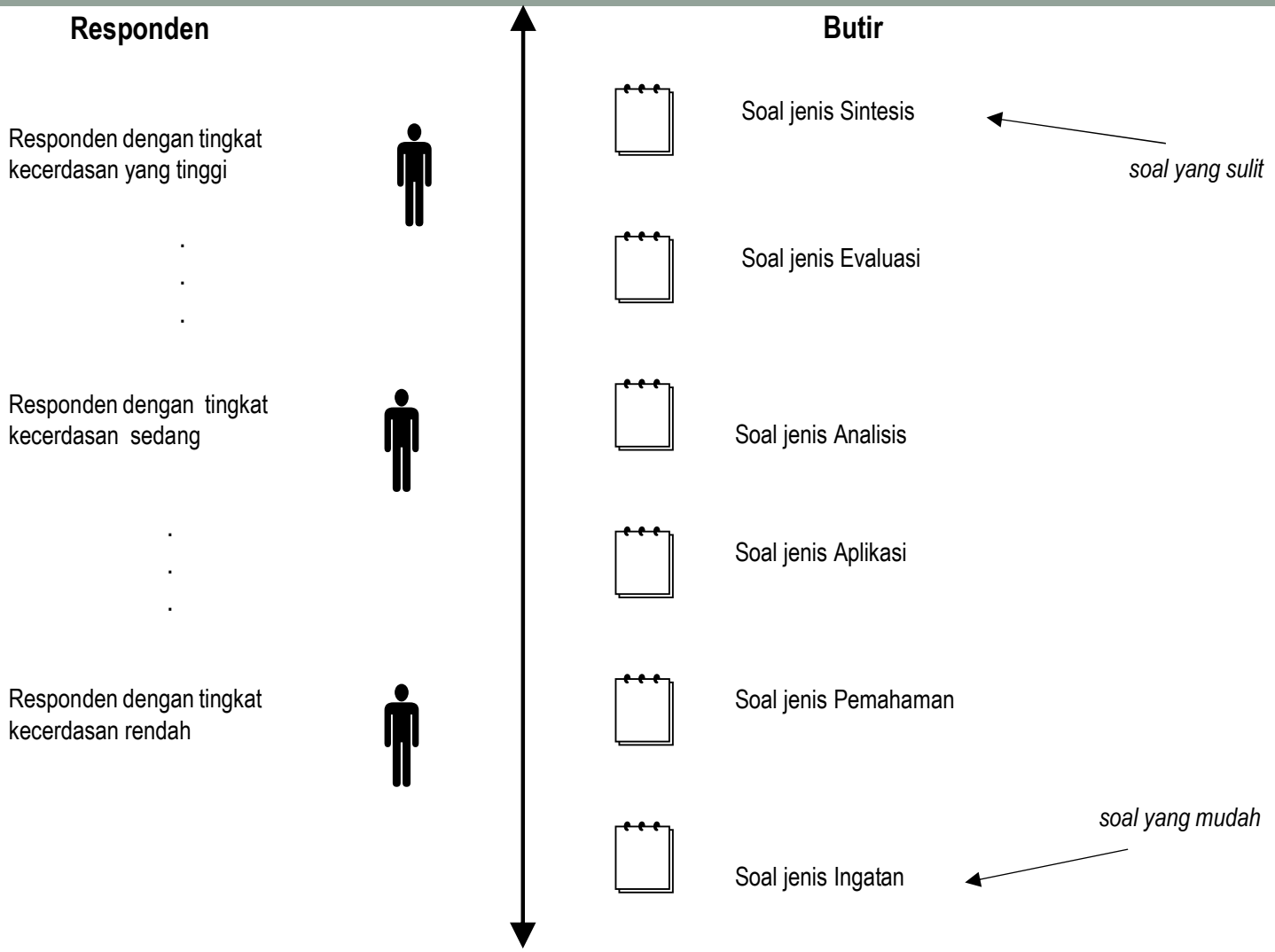
Tingkat Kesulitan Tinggi

Butir

Tingkat Kesulitan Rendah



Philadelphia Naming Test (PNT; Roach, Schwartz, Martin, Grewal, & Brecher, 1996)

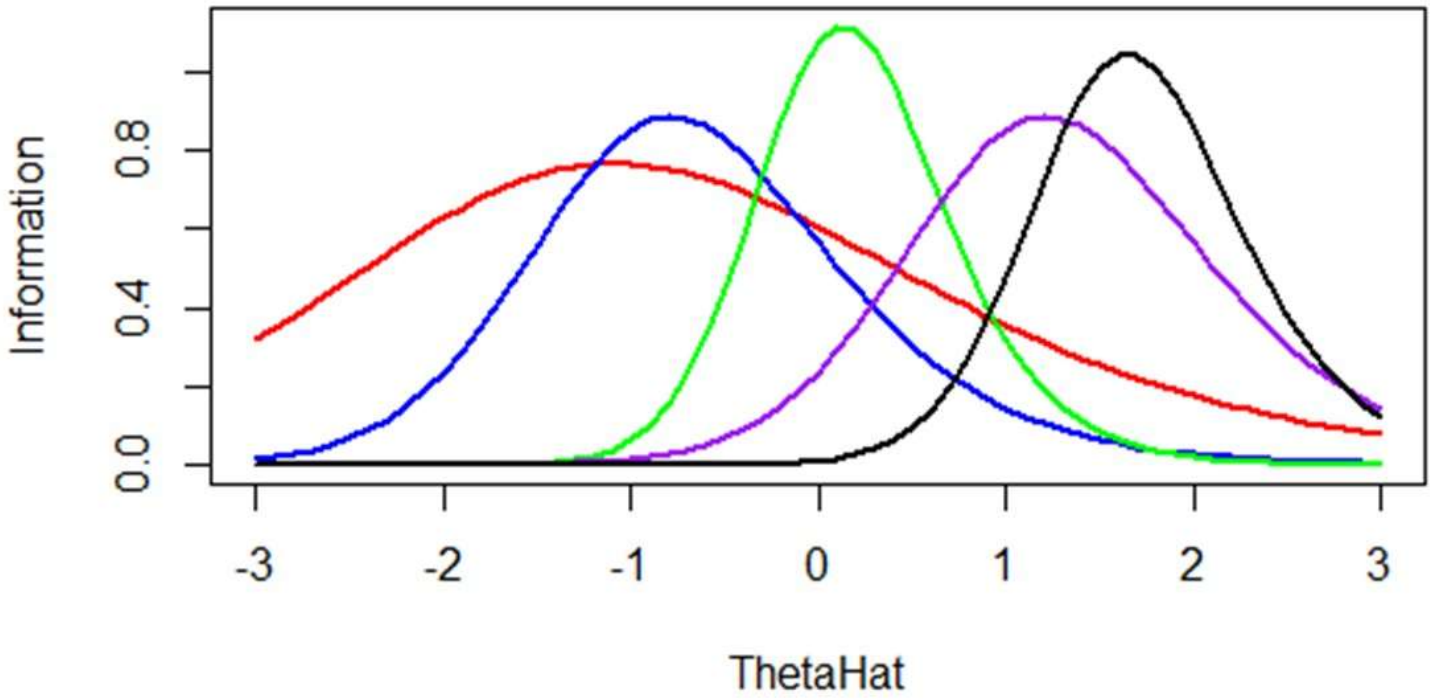

Responden
Butir
Responden dengan tingkat kecerdasan yang tinggi
Responden dengan tingkat kecerdasan sedang
Responden dengan tingkat kecerdasan rendah
Soal jenis Sintesis
soal yang sulit
Soal jenis Evaluasi
Soal jenis Analisis
Soal jenis Aplikasi
Soal jenis Pemahaman
soal yang mudah
Soal jenis Ingatan

Properti Tes

- **Reliabilitas**
  - Person reliability (>.80)
  - Item reliability (> .90)
- **Separation**
  - Person Separation (G) (> 2)
  - Item Separation (G) (> 3)
- **Peta Butir Orang (*item person map*)**
  - Sebaran skor individu harus match dengan sebaran butir
  - Setiap interval terisi dengan butir
  - Apakah sesuai dengan tujuan tes? (seleksi, remidi)

Lain-lain

- Butir yang sulit atau mudah memiliki daya disriminasi rendah
- Butir memiliki tingkat kesulitan sangat tinggi (di atas sebagian rerata orang) biasanya merupakan soal salah kunci



Manakah profil tes yang cocok untuk seleksi?
Manakah profil tes yang cocok untuk remidi?
Manakah profil tes yang cocok untuk identifikasi kemampuan moderat

PARAMETER BUTIR

PARAMETER BUTIR

PARAMETER BUTIR

Parameter Butir

Skrining Awal

- **Kesalahan kunci**
  - Item measure – (EMPIRICAL ITEM CATEGORY MEASURE)

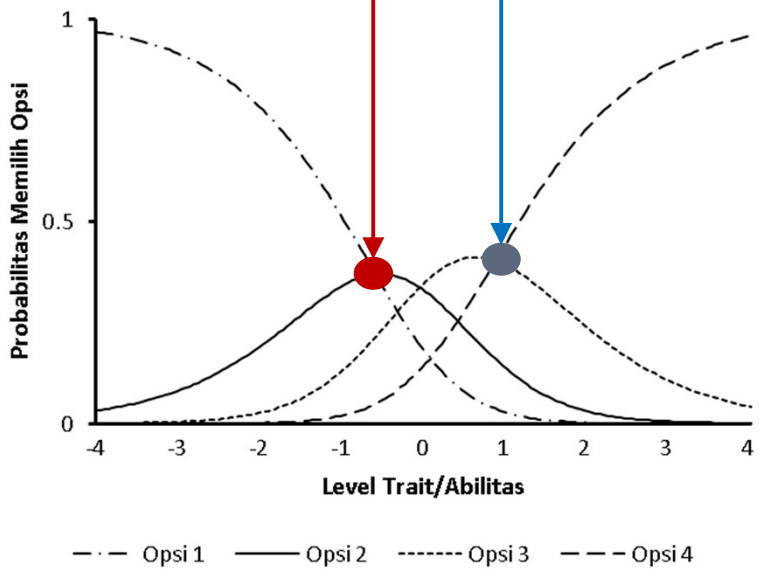
Properti Rasch

- **Tingkat kesulitan Butir (MEASURE)**
- **Ketepatan Model Item (INFIT/OUTFIT)**

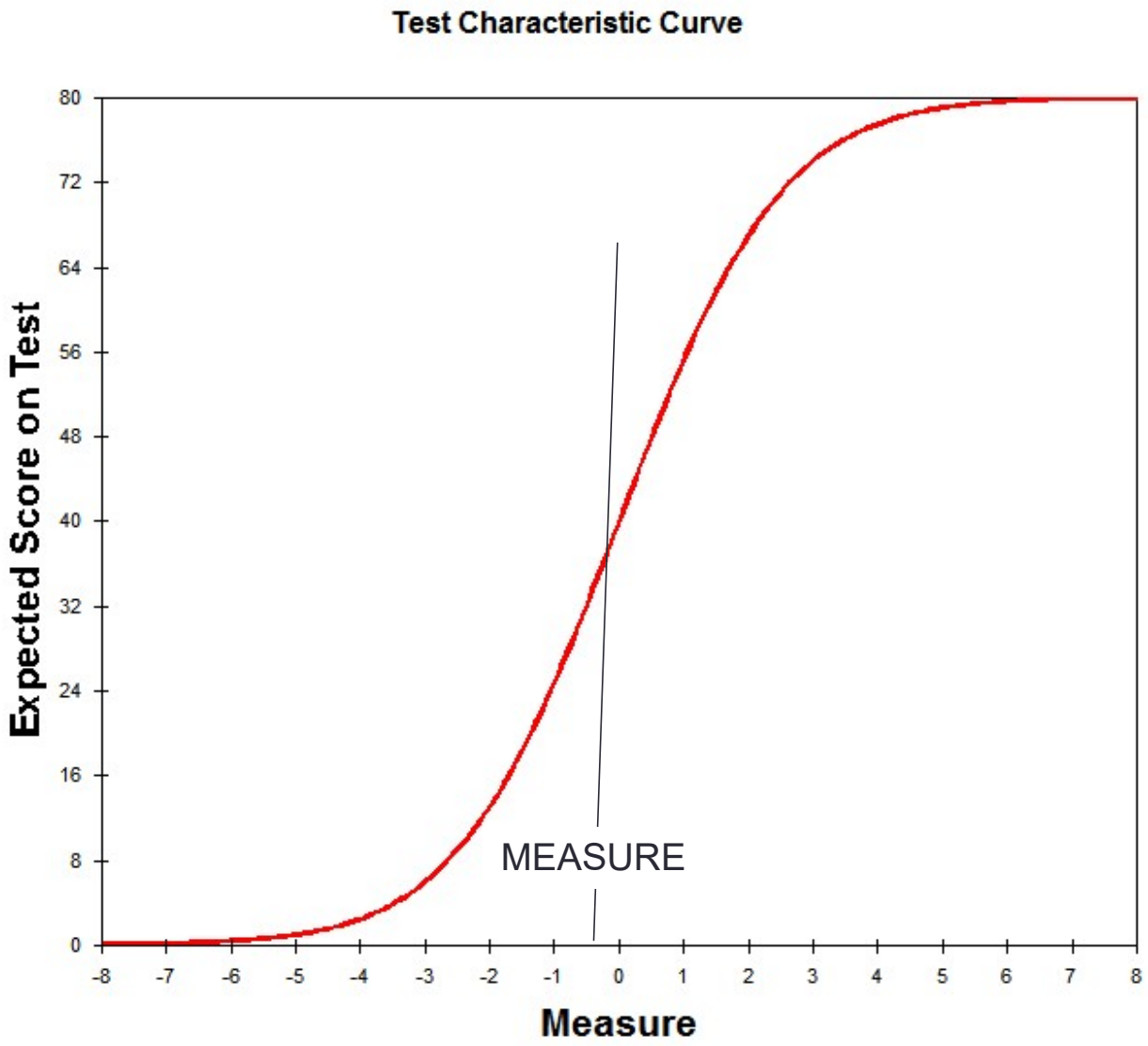
OUTPUT WINSTEP : Measure

TABLE 3.2 (Klik Output 3.2 Rating (Partial Credit) Scale

CATEGORY LABEL	STRUCTURE MEASURE	S.E.	SCORE-TO-MEASURE AT CAT.	----ZONE----		50% CUM. PROBABLTY	COHERENCE M->C	C->M	RMSR	ESTIM DISCR
1	NONE		(-1.97)	-INF	-1.27		0%	0%	1.3334	
2	-.57	.12	-.49	-1.27	.08	-.94	28%	36%	.6697	1.29
3	-.02	.09	.67	.08	1.50	.06	42%	74%	.4346	1.00
4	.91	.10	(2.25)	1.50	+INF	1.21	83%	7%	1.2563	.81



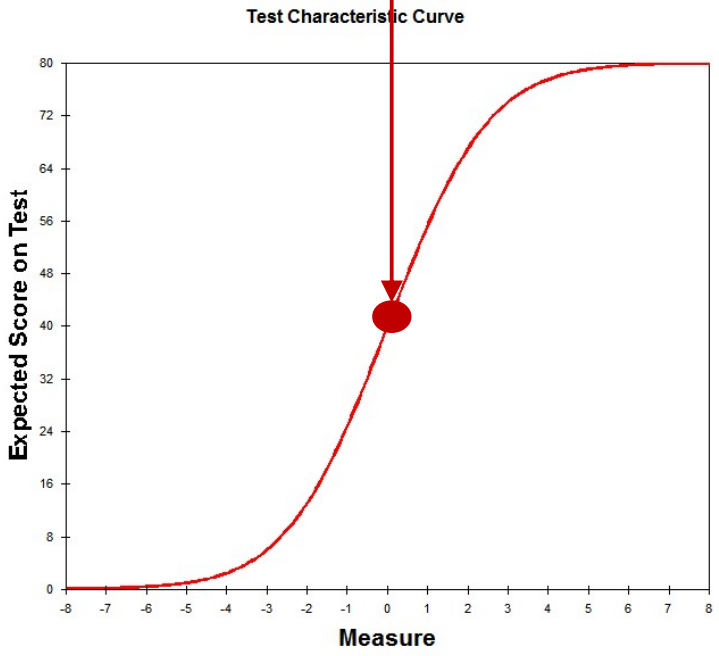
Measure menunjukkan nilai lokasi butir, yaitu perpotongan antara probabilitas dua kategori bawah dan atasnya


Test Characteristic Curve
Expected Score on Test
80
72
64
56
48
40
32
24
16
8
0
-8
-7
-6
-5
-4
-3
-2
-1
0
1
2
3
4
5
6
7
8
Measure
MEASURE

OUTPUT WINSTEP : Measure

TABLE 3.2 (Klik Output 3.2 Rating (Partial Credit) Scale

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PTBISERL-EX CORR.	PTBISERL-EX EXP.	EXACT MATCH OBS%	EXACT MATCH EXP%	ITEM	G
1	7	40	.80	.43	1.06	.3	1.21	.7	.04	.18	82.5	82.5	I0001	0
2	1	40	2.96	1.02	1.05	.4	3.92	1.8	-.27	.07	97.5	97.5	I0002	0
3	6	37	.96	.45	1.03	.2	1.33	.9	.05	.15	83.8	83.8	I0003	0

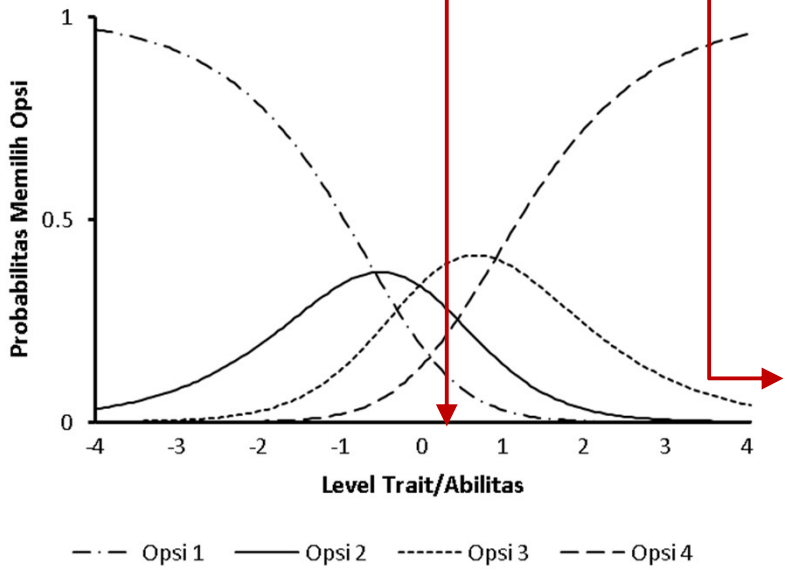


MEASURE menunjukkan nilai lokasi butir, yaitu perpotongan antara probabilitas dua kategori bawah dan atasnya

OUTPUT WINSTEP : Kecocokan Model

TABLE 14.1 (Klik Output Table 14.1)

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PT-MEASURE CORR.	PT-MEASURE EXP.
1	1649	625	.11	.04	.99	-.2	.99	-.1	.41	.40



Measure

Measure menunjukkan tingkat kesulitan butir.

Contoh butir 1 memiliki tingkat kesulitan sedang, karena nilai tingkat kesulitan butir sebesar 0.11 (mendekati 0)

MNSQ (Indeks Ketepatan Model)

>1.5 : terlalu banyak eror

0.5-1.5 : fit/produktif

<0.5: overfit

OUTPUT WINSTEP : Uji DIF

TABLE 14.1 (Klik Output Table 14.1)

PERSON CLASSES	SUMMARY DIF CHI-SQUARE	D.F.	PROB.	BETWEEN-CLASS MEAN-SQUARE	t=ZSTD	ITEM Number	Name
2	7.7960	1	.0052	3.8446	1.6733	1	A1
2	70.1075	1	.0000	35.1189	5.2969	2	A6
2	10.9977	1	.0009	5.4713	2.0881	3	A8
2	66.4634	1	.0000	32.3398	5.1086	4	A10
2	.0000	1	1.0000	.0052	-1.2834	5	A11
2	4.6518	1	.0310	2.3220	1.1591	6	A13

Terjangkit (*Infected*) DIF ?
P<0.05 : terjangkit DIF
P>0.05 : tidak terjangkit DIF

Butir 10 tidak terjangkit DIF

OUTPUT WINSTEP : Uji DIF

TABLE 14.1 (Klik Output Table 14.1)

PERSON CLASS	DIF MEASURE	DIF S.E.	PERSON CLASS	DIF MEASURE	DIF S.E.	DIF CONTRAST	ITEM Number	Item Name
1	1.14	.06	2	.91	.05	.23	1	A1
1	.24	.06	2	.90	.05	-.66	2	A6
1	-1.22	.05	2	-.98	.05	-.24	3	A8
1	1.44	.07	2	.70	.05	.73	4	A10
1	-1.60	.07	2	-1.60	.06	.00	5	A11
1	.23	.07	2	.03	.06	.20	6	A13
2	.91	.05	1	1.14	.06	-.23	1	A1
2	.90	.05	1	.24	.06	.66	2	A6

Tingkat Kesukaran Butir Berdasarkan Jenis Kelamin (Person Class)
Terlihat bahwa tingkat kesukaran butir pada laki-laki (kode 1; b= 1.14) lebih tinggi dibanding dengan wanita (kode=2; 0.91)

OUTPUT WINSTEP : Person Fit

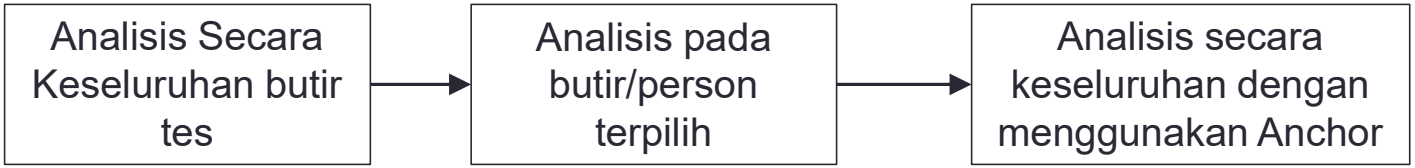
TABLE 14.1 (Klik Output Table 14.1)

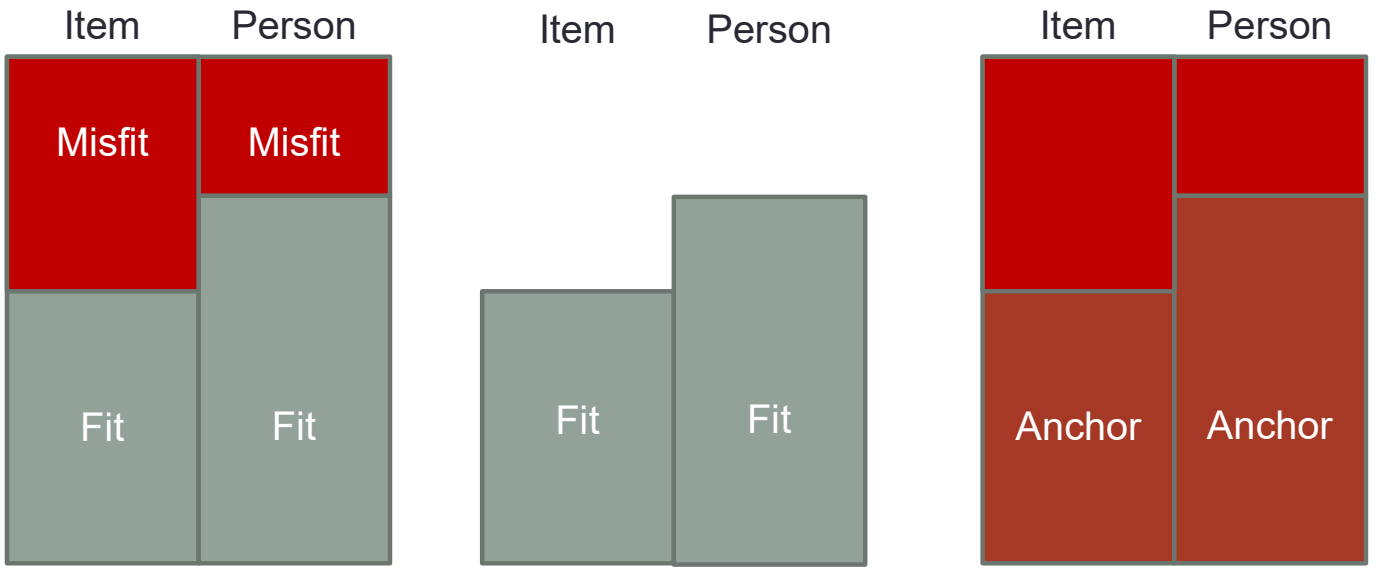
ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PT-MEASURE CORR.	PT-MEASURE EXP.	EXACT MATCH OBS%	EXACT MATCH EXP%	PERSON
1	2	6	-3.07	1.03	.57	-.7	.39	-.4	.52	.51	83.3	76.4	1
2	3	6	-2.11	.95	1.07	.3	.95	.2	.70	.54	66.7	76.8	1
3	5	6	-.49	.85	1.10	.4	.91	.1	.64	.46	66.7	71.4	1
4	4	6	-1.25	.90	1.68	1.1	1.48	.8	-.18	.50	50.0	72.4	1
5	6	6	.20	.82	1.88	1.4	1.69	1.0	.50	.42	33.3	69.0	1
6	8	6	1.49	.80	1.83	1.4	1.78	1.3	.12	.36	33.3	67.6	1
7	5	6	-.49	.85	.31	-1.4	.26	-1.2	.43	.46	100.0	71.4	1
8	4	6	-1.25	.90	.42	-1.0	.35	-1.0	.70	.50	83.3	72.4	1

MNSQ (Indeks Ketepatan Model)
>1.5 : terlalu banyak eror
0.5-1.5 : fit/produktif
<0.5: overfit

ISU-ISU TAMBAHAN

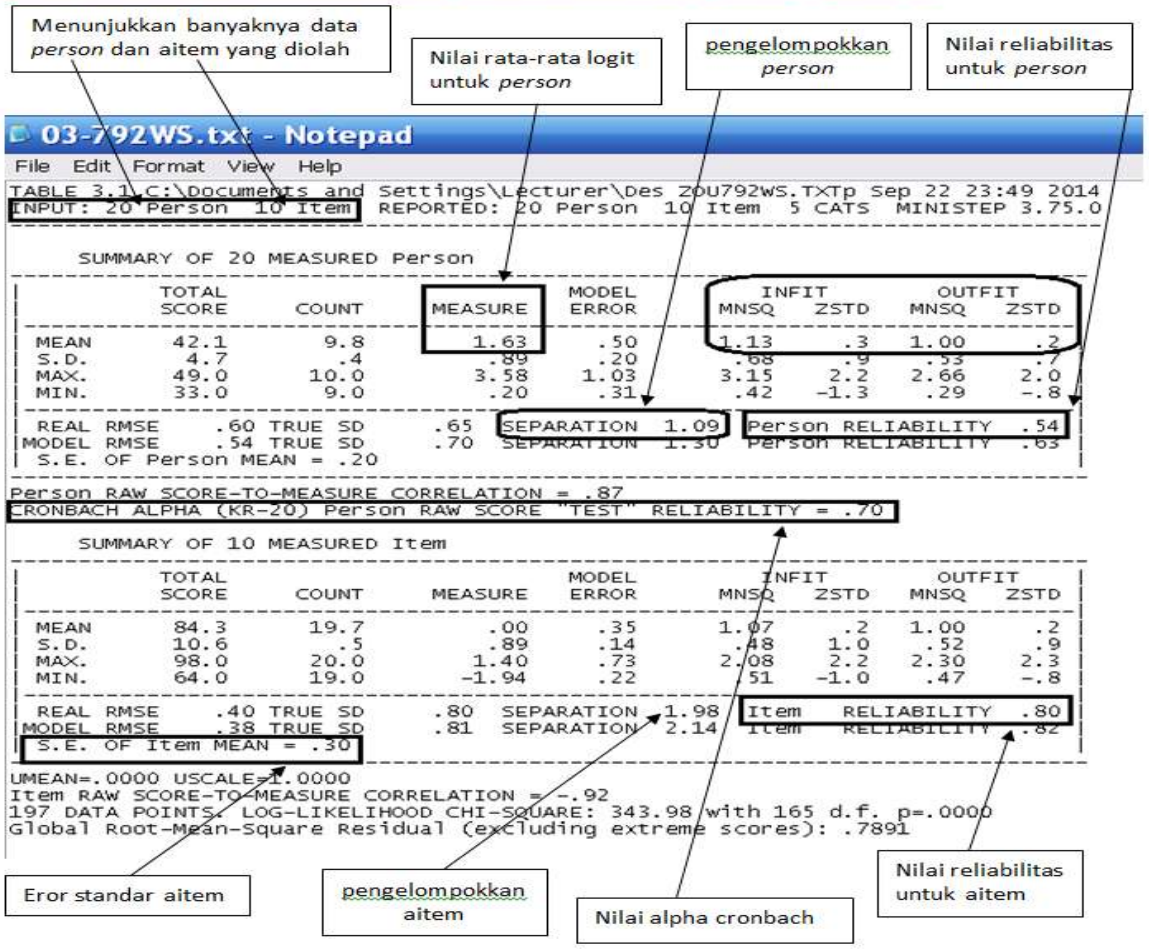
Prosedur Analisis


Analisis Secara Keseluruhan butir tes
Analisis pada butir/person terpilih
Analisis secara keseluruhan dengan menggunakan Anchor



Winsteps: ringkasan statistik

Tabel ringkasan statistik untuk berkas ***latihanWinsteps.prn*** seperti di bawah ini:


Menunjukkan banyaknya data *person* dan aitem yang diolah
Nilai rata-rata logit untuk *person*
pengelompokkan *person*
Nilai reliabilitas untuk *person*
03-792WS.txt - Notepad
File Edit Format View Help
TABLE 3.1 C:\Documents and Settings\Lecturer\Des ZOU792WS.TXTp Sep 22 23:49 2014
INPUT: 20 Person 10 Item REPORTED: 20 Person 10 Item 5 CATS MINISTEP 3.75.0
SUMMARY OF 20 MEASURED Person
TOTAL SCORE COUNT MEASURE MODEL ERROR INFIT MNSQ ZSTD OUTFIT MNSQ ZSTD
MEAN 42.1 9.8 1.63 .50 1.13 .3 1.00 .2
S.D. 4.7 .4 .89 .20 .68 .9 .53 .7
MAX. 49.0 10.0 3.58 1.03 3.15 2.2 2.66 2.0
MIN. 33.0 9.0 .20 .31 .42 -1.3 .29 -.8
REAL RMSE .60 TRUE SD .65 SEPARATION 1.09 Person RELIABILITY .54
MODEL RMSE .54 TRUE SD .70 SEPARATION 1.30 Person RELIABILITY .63
S.E. OF Person MEAN = .20
Person RAW SCORE-TO-MEASURE CORRELATION = .87
CRONBACH ALPHA (KR-20) Person RAW SCORE "TEST" RELIABILITY = .70
SUMMARY OF 10 MEASURED Item
TOTAL SCORE COUNT MEASURE MODEL ERROR INFIT MNSQ ZSTD OUTFIT MNSQ ZSTD
MEAN 84.3 19.7 .00 .35 1.07 .2 1.00 .2
S.D. 10.6 .5 .89 .14 .48 1.0 .52 .9
MAX. 98.0 20.0 1.40 .73 2.08 2.2 2.30 2.3
MIN. 64.0 19.0 -1.94 .22 .51 -1.0 .47 -.8
REAL RMSE .40 TRUE SD .80 SEPARATION 1.98 Item RELIABILITY .80
MODEL RMSE .38 TRUE SD .81 SEPARATION 2.14 Item RELIABILITY .82
S.E. OF Item MEAN = .30
UMEAN=.0000 USCALE=1.0000
Item RAW SCORE-TO-MEASURE CORRELATION = -.92
197 DATA POINTS, LOG-LIKELIHOOD CHI-SQUARE: 343.98 with 165 d.f. p=.0000
Global Root-Mean-Square Residual (excluding extreme scores): .7891
Eror standar aitem
pengelompokkan aitem
Nilai alpha cronbach
Nilai reliabilitas untuk aitem

Winsteps: analisis aitem/butir

Nilai logit untuk masing-masing aitem diurutkan dari tertinggi ke terendah

13-792WS.txt - Notepad

File Edit Format View Help

TABLE 13.1 C:\Documents and Settings\Lecturer\De ZOU792WS.TXT. Sep 22 23:49 2014
INPUT: 20 Person 10 Item REPORTED: 20 Person 10 Item 5 CATS MINISTEP 3.75.0

Person: REAL SEP.: 1.09 REL.: .54 ... Item: REAL SEP.: 1.98 REL.: .80

Item STATISTICS: MEASURE ORDER

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PT-MEASURE CORR.	PT-MEASURE EXP.	EXACT MATCH OBS%	EXACT MATCH EXP%	Item
3	66	20	1.40	.22	.79	-.6	.83	-.4	.62	.67	50.0	36.5	N3
2	64	19	1.27	.23	1.28	.9	1.32	1.0	.50	.64	31.6	36.6	N2
5	81	19	.26	.29	1.65	1.5	2.30	2.3	.25	.51	63.2	55.5	N5
1	86	20	.22	.29	2.08	2.2	1.30	.8	.61	.50	60.0	55.0	N1
10	87	20	.13	.30	.97	.1	1.01	.2	.54	.48	65.0	57.2	N10
9	90	20	-.17	.34	.76	-.4	.63	-.7	.53	.44	65.0	64.2	N9
7	91	20	-.29	.36	1.30	.7	.95	.1	.46	.42	65.0	65.4	N7
6	87	19	-.32	.38	.51	-1.0	.53	-.8	.61	.42	68.4	67.1	N6
4	93	20	-.57	.40	.55	-.9	.67	-.5	.50	.38	85.0	69.1	N4
8	98	20	-1.94	.73	.83	.0	.47	-.3	.37	.22	90.0	90.6	N8
MEAN	84.3	19.7	.00	.35	1.07	.2	1.00	.2			64.3	59.7	
S.D.	10.6	.5	.89	.14	.48	1.0	.52	.9			15.5	15.0	

Jumlah data '19', menunjukkan adanya data yang hilang

Nilai devisasi standar logit item berguna untuk identifikasi kelompok aitem (separation)

Nilai rata-rata aitem selalu 0,00 logit

Label aitem

Winsteps: analisis person/responden

Nilai logit untuk masing-masing *person* diurutkan dari tertinggi ke terendah

Label *person*

17-792WS.txt - Notepad

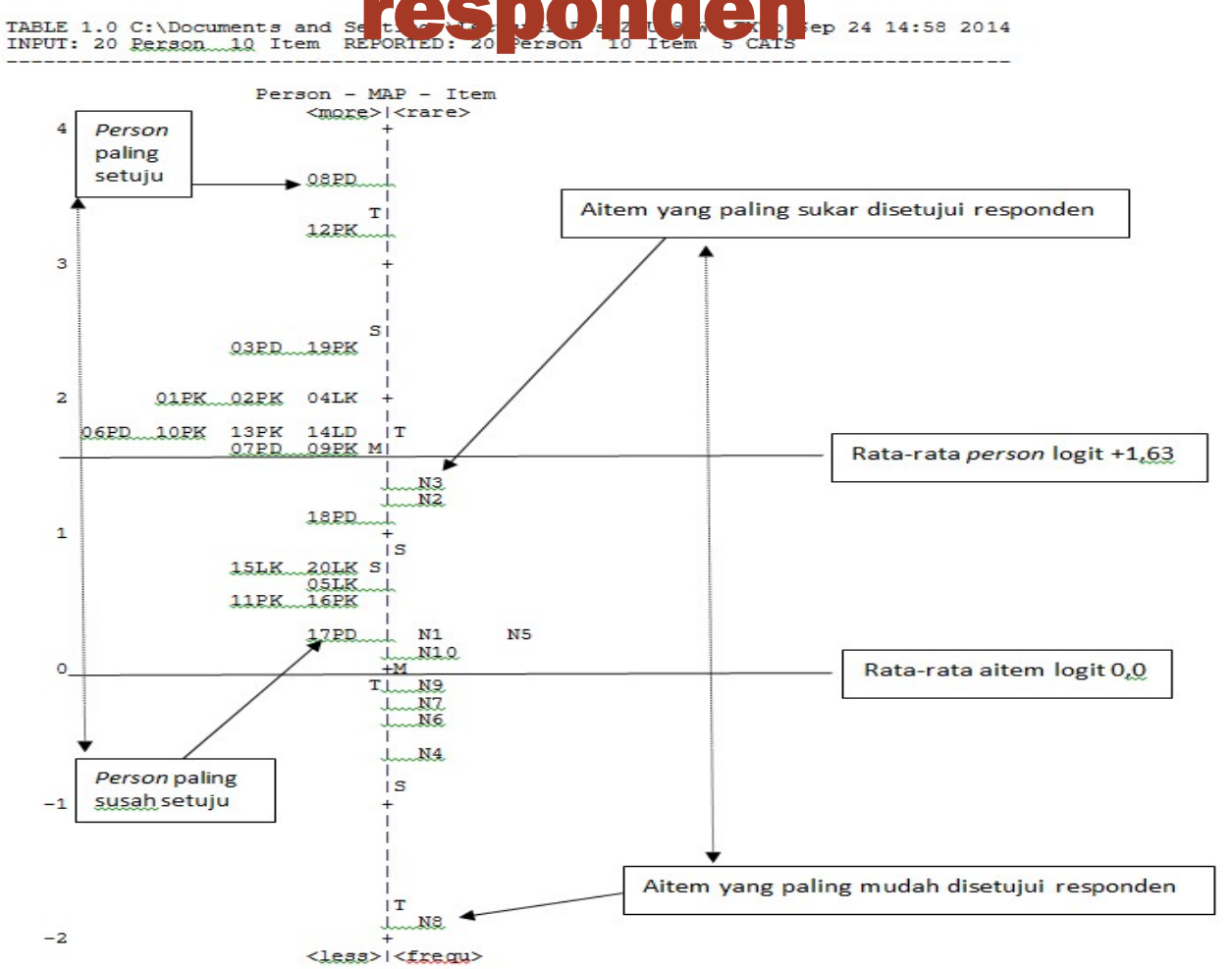
File Edit Format View Help

```
TABLE 17.1 C:\Documents and Settings\Lecturer\De ZOU792WS.TXT. Sep 22 23:49 2014
INPUT: 20 Person  10 Item  REPORTED: 20 Person  10 Item  5 CATS  MINISTEP 3.75.0
-------------------------------------------------------------------------------
Person: REAL SEP.: 1.09  REL.: .54 ... Item: REAL SEP.: 1.98  REL.: .80

         Person STATISTICS:  MEASURE ORDER
```

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PT-MEASURE CORR.	PT-MEASURE EXP.	EXACT MATCH OBS%	EXACT MATCH EXP%	Person
8	49	10	3.58	1.03	.66	.0	.29	-.2	.52	.23	90.0	90.6	08PD
12	44	9	3.28	1.03	1.03	.4	.95	.4	.12	.24	88.9	89.6	12PK
3	47	10	2.38	.61	.76	-.1	1.00	.3	.27	.37	70.0	74.2	03PD
19	47	10	2.38	.61	1.39	.7	1.35	.7	.34	.37	80.0	74.2	19PK
1	46	10	2.06	.53	1.26	.6	.66	-.2	.67	.41	90.0	70.3	01PK
2	46	10	2.06	.53	.64	-.4	.57	-.4	.64	.41	80.0	70.3	02PK
4	46	10	2.06	.53	3.15	2.2	1.53	.9	.47	.41	80.0	70.3	04LK
10	45	10	1.81	.48	.51	-.7	.62	-.4	.60	.45	70.0	66.0	10PK
13	45	10	1.81	.48	1.05	.3	.62	-.4	.74	.45	80.0	66.0	13PK
14	45	10	1.81	.48	.44	-.9	.63	-.4	.52	.45	70.0	66.0	14LD
6	40	9	1.69	.48	.54	-.6	.83	.0	.39	.49	55.6	62.0	06PD
7	44	10	1.60	.44	2.54	2.0	2.66	2.0	.35	.48	60.0	54.6	07PD
9	44	10	1.60	.44	.42	-1.1	.45	-.8	.74	.48	60.0	54.6	09PK
18	41	10	1.11	.37	1.28	.7	.96	.1	.58	.56	40.0	44.3	18PD
15	34	9	.77	.36	1.49	1.0	1.88	1.4	-.12	.61	55.6	41.7	15LK
20	38	10	.73	.34	1.23	.6	1.14	.4	.63	.61	50.0	41.7	20LK
5	37	10	.62	.33	1.55	1.1	1.25	.6	.49	.62	50.0	41.7	05LK
11	36	10	.51	.33	.46	-1.3	.69	-.5	.60	.63	50.0	41.4	11PK
16	36	10	.51	.33	1.01	.2	.90	.0	.51	.63	40.0	41.4	16PK
17	33	10	.20	.31	1.16	.5	1.00	.2	.74	.65	30.0	36.7	17PD
MEAN	42.1	9.8	1.63	.50	1.13	.3	1.00	.2			64.5	59.9	
S.D.	4.7	.4	.89	.20	.68	.9	.53	.7			17.6	16.1	

Winsteps: peta aitem-responden



Winsteps: unidimensionalitas

23-792WS.txt - Notepad

File Edit Format View Help

```
TABLE 23.0 C:\Documents and Settings\Lecturer\De ZOU792WS.TXT. Sep 22 23:49 2014
INPUT: 20 Person  10 Item  REPORTED: 20 Person  10 Item  5 CATS  MINISTEP 3.75.0
-------------------------------------------------------------------------------

     Table of STANDARDIZED RESIDUAL variance (in Eigenvalue units)
                                             -- Empirical --     Modeled
Total raw variance in observations     =         17.4 100.0%          100.0%
  Raw variance explained by measures   =          7.4  42.6%           49.6%
    Raw variance explained by persons  =          3.0  17.4%           20.2%
    Raw variance explained by items    =          4.4  25.2%           29.4%
  Raw unexplained variance (total)     =         10.0  57.4% 100.0%    50.4%
    Unexplned variance in 1st contrast =          2.3  13.5%  23.4%
    Unexplned variance in 2nd contrast =          1.9  10.7%  18.6%
    Unexplned variance in 3rd contrast =          1.4   8.1%  14.0%
    Unexplned variance in 4th contrast =          1.2   6.9%  12.1%
    Unexplned variance in 5th contrast =          1.0   5.7%   9.9%
```

Winsteps: pengujian skala peringkat

41-792WS.txt - Notepad

File Edit Format View Help

TABLE 3.2 C:\Documents and Settings\Lecturer\Des ZOU792WS.TXTp Sep 22 23:49 2014
INPUT: 20 Person 10 Item REPORTED: 20 Person 10 Item 5 CATS MINISTEP 3.75.0

SUMMARY OF CATEGORY STRUCTURE. Model="R"

CATEGORY LABEL	SCORE	OBSERVED COUNT	%	OBSVD AVRGE	SAMPLE EXPECT	INFIT MNSQ	OUTFIT MNSQ	ANDRICH THRESHOLD	CATEGORY MEASURE	
1	1	8	4	.08	-.44	1.94	2.52	NONE	(-1.86)	1
2	2	7	4	.02*	.01	1.07	.99	-.10	-.81	2
3	3	18	9	.33	.56	.82	.63	-.67	-.13	3
4	4	53	27	1.10	1.24	1.15	.79	-.20	.69	4
5	5	111	56	2.31	2.24	.76	.88	.97	(2.26)	5
MISSING		3	2	1.51						

OBSERVED AVERAGE is mean of measures in category. It is not a parameter estimate.

PENULISAN HASIL

Table 3: Individual item fit statistics for the final models of the revised DASS-21 subscales

Item no.	Item	Location value	SE	Fit Resid	ChiSq	Prob
	DASS-21-Depression					
3	I couldn't seem to experience any positive feeling at all	0.059	0.101	0.742	3.107	0.54
10	I felt that I had nothing to look forward to	0.048	0.097	-1.477	8.234	0.08
13	I felt down-hearted and blue	-1.100	0.089	0.685	5.280	0.26
16	I was unable to become enthusiastic about anything	-0.120	0.097	-0.284	8.884	0.06
17	I felt I wasn't worth much as a person	0.296	0.099	-0.248	3.719	0.45
21	I felt that life was meaningless	0.817	0.109	-0.680	9.191	0.06
	DASS-21-Anxiety					
4	I experienced breathing difficulty	0.115	0.086	-0.210	7.169	0.13
7	I experienced trembling (e.g. in the hands)	0.108	0.086	-0.413	5.858	0.21
9	I was worried about situations in which I might panic and make a fool of myself	-0.483	0.077	0.361	2.220	0.70
15	I felt I was close to panic	0.061	0.087	-1.836	13.075	0.01
19	I was aware of the action of my heart in the absence of physical exertion	-0.100	0.083	0.304	5.257	0.26
20	I felt scared without any good reason	0.298	0.090	-1.311	6.987	0.14
	DASS-21-Stress					
1	I found it hard to wind down	-0.552	0.070	0.600	6.544	0.16
6	I tended to over-react to situations	0.001	0.070	0.313	3.490	0.48
8	I felt that I was using a lot of nervous energy	0.313	0.070	0.107	3.117	0.54
12	I found it difficult to relax	-0.145	0.072	-2.648	14.029	0.007
14	I was intolerant of anything that kept me from getting on with what I was doing	0.346	0.074	1.059	4.195	0.38
18	I felt that I was rather touchy	0.038	0.072	-1.033	10.712	0.03

SE = Standard error, Fit Resid = Fit Residual, ChiSq = Chi-square, prob = probability
Anxiety: Fit Residual df = 208, Chi square df = 4